CALIFORNIA
OPEN DATA
HANDBOOK

DRAFT - JANUARY, 7 2014

Stewards of Change
INSTITUTE, INC

Content

This handbook was compiled from various sources, but the majority of the information comes from four primary sources: The Open Knowledge Foundation, The Sunlight Foundation, The World Bank and The New York Open Data Portal, all of which are respected and trusted leaders in this sector. The World Bank: The World Bank authorizes the use of this material subject to the terms and conditions on its website, http://www.worldbank.org/terms Likewise, The Open Knowledge Foundation and The Sunlight Foundation operate under the Creative Commons Attribution 3.0 United States License and the individual credits are enumerated in the appropriate sections below.

While the success of commercial products is generally measured by sales and revenue numbers, Open Data's success is more accurately measured by its use and reuse with the often unspoken desire that it become viral and serves the greater good by positively impacting the world. The motivation for someone to voluntarily contribute their data for free without any restrictions is generally rooted in the passion and quest to make a meaningful contribution to society. And so, it almost seems as if it is our civic duty to take the torch and pass it along in the hope that the next person will be inspired to do the same, thereby honoring the original intentions and furthering the ultimate cause. It is in this context and with that intention, then, that we have assembled the following open data about open data. For, had it not been for the many people that contributed this information the way they did, this handbook most likely would not have been created, and likewise the resulting ripple effect which can forever change the world in which we live

The handbook discusses the legal, social and technical aspects of open data. It can be used by anyone but is especially designed for those seeking to open up data. It discusses the why, what and how of open data – why to go open, what open is, and the how to "open" data.

Note, the printed version of this document in its present form does not contain the appendices that are included in the electronic version. In the future, it is our hope that the electronic version will become a wiki with additions and refinements contributed by its readers.

At present, you can find the full document including appendices, as a pdf file, at the Open Datafest Collaboration Site or by emailing mkerr@stewardsofchange.com.

Open Data

1.1 Introduction

Do you know exactly how much of your tax money is spent on street lights or on cancer research? What is the shortest, safest and most scenic bicycle route from your home to your work? And what is in the air that you breathe along the way? Where in your region will you find the best job opportunities and the highest number of fruit trees per capita? When can you influence decisions about topics you deeply care about, and whom should you talk to?

New technologies now make it possible to build the services to answer these questions automatically. Much of the data you would need to answer these questions is generated by public bodies. However, often the data required is not yet available in a form which is easy to use. This book is about how to unlock the potential of official and other information to enable new services, to improve the lives of citizens and to make government and society work better.

The notion of _open data_ and specifically _open government data_ - information, public or otherwise, which anyone is free to access and re-use for any purpose - has been around for some years. In 2009, open data started to become visible in the mainstream, with various governments (such as the USA, UK, Canada and New Zealand) announcing new initiatives towards opening up their public information.

This book explains the basic concepts of 'open data', especially in relation to government. It covers how open data creates value and can have a positive impact in many different areas. In addition to exploring the background, the handbook also provides concrete information on how to produce open data.

1.1.2 Target Audience

This handbook has a broad audience:

- for those who have never heard of open data before and those who consider themselves seasoned data professionals
- for civil servants and for activists
- for journalists and researchers
- for politicians and developers
- for data geeks and those who have never heard of an API.

Most of the information currently provided is focused on data held by the public sector. However, the author's intentions are to broaden this as time permits. You are welcome to participate to help us with that effort.

This handbook is intended to be easily understood by those with little or no knowledge of the topic. If you do find a piece of jargon or terminology with which you aren't familiar, please see the detailed Glossary and FAQs (frequently asked questions) which can be found at the end of the handbook.

1.1.3 Credits

Credits and Copyright

Contributing authors

- Daniel Dietrich
- Jonathan Gray
- Tim McNamara
- Antti Poikola
- Rufus Pollock
- Julian Tait
- Ton Zijlstra

Existing sources directly used

- Technical Proposal for how IATI is implemented. *The IATI Technical Advisory Group led by Simon Parrish*
- Unlocking the Potential of Aid Information. *Rufus Pollock, Jonathan Gray, Simon Parrish, Jordan Hatcher*
- Finnish manual authored by *Antti Poikola*
- Beyond Access Report. *Access Info and the Open Knowledge Foundation*

Other sources

- W3C Publishing Government Data (2009) http://www.w3.org/TR/gov-data/

1.2 Why Open Data?

Open data, especially open government data, is a tremendous resource that is as yet largely untapped. Many individuals and organizations collect a broad range of different types of data in order to perform their tasks. Government is particularly significant in this respect, both because of the quantity and centrality of the data it collects, but also because most of that government data is public data by law, and therefore could be made open and made available for others to use. Why is that of interest?

There are many areas where we can expect open data to be of value, and where examples of how it has been used already exist. There are also many different groups of people and organizations who can benefit from the availability of open data, including government itself. At the same time it is impossible to predict precisely how and where value will be created in the future. The nature of innovation is that developments often come from unlikely places.

It is already possible to point to a large number of areas where open government data is creating value. Some of these areas include:

• Transparency and democratic control

• Participation

• Self-empowerment

• Improved or new private products and services

• Innovation

• Improved efficiency of government services

• Improved effectiveness of government services

• Impact measurement of policies

• New knowledge from combined data sources and patterns in large data volumes Examples exist for most of these areas.

In terms of transparency, projects such as the Finnish 'tax tree' and British 'where does my money go' show how your tax money is being spent by the government. And there's the example of how open data saved Canada $3.2 billion in charity tax fraud. Also various websites such as the Danish folketsting.dk track activity in parliament and the law making processes, so you can see what exactly is happening, and which parliamentarians are involved.

Open government data can also help you to make better decisions in your own life, or enable you to be more active in society.

 A woman in Denmark built findtoilet.dk, which showed all the Danish public toilets, so that people she knew with bladder problems can now trust themselves to go out more again. In the Netherlands a service, vervuilingsalarm.nl, is available which warns you with a message if the air-quality in your vicinity is going to reach a self-defined threshold tomorrow. In New York you can easily find out where you can walk your dog, as well as find other people who use the same parks. Services like 'mapumental' in the UK and 'mapnificent' in Germany allow you to find places to live, taking into account the duration of your commute to work, housing prices, and how beautiful an area is. All these examples use open government data.

Economically, open data is of great importance as well. Several studies have estimated the economic value of open data at several tens of billions of Euros annually in the EU alone. New products and companies are re-using open data. The Danish husetsweb.dk helps you to find ways of improving the energy efficiency of your home, including financial planning and finding builders who can do the work. It is based on re-using cadastral information and information about government subsidies, as well as the local trade register. Google Translate uses the enormous volume of EU documents that appear in all European languages to train the translation algorithms, thus improving its quality of service.

Open data is also of value for government itself. For example, it can increase government efficiency. The Dutch Ministry of Education has published all of its education-related data online for re-use. Since then, the number of questions they receive has dropped, reducing work-load and costs, and the remaining questions are now also easier for civil servants to answer, because it is clear where the relevant data can be found. Open data is also making government more effective, which ultimately also reduces costs. The Dutch department for cultural heritage is actively releasing its data and collaborating with amateur historical societies and groups such as the Wikimedia Foundation in order to execute their own tasks more effectively. This not only results in improvements to the quality of their data, but will also ultimately make the department smaller.

While there are numerous instances of the ways in which open data is already creating both social and economic value, we don't yet know what new things will become possible.

New combinations of data can create new knowledge and insights, which can lead to whole new fields of application.

We have seen this in the past, for example when Dr. Snow discovered the relationship between drinking water pollution and cholera in London in the 19th century, by combining data about cholera deaths with the location of water wells. This led to the building of London's sewage systems, and hugely improved the general health of the population. We are likely to see such developments happening again as unexpected insights flow from the combination of different open data sets.

The following are some studies that explain and try to quantify these benefits:

http://wiki.linkedgov.org/index.php/The_economic_impact_of_open_data
(The Economic Impact of Open Data)

http://blog.opengovpartnership.org/2012/07/open-data-economic-growth/
(Open Data and Economic Growth)

http://ands.org.au/resource/cost-benefit.html

(Costs and Benefits of Data Provision)

http://openeconomics.net/2012/10/03/the-benefits-of-open-data-evidence-from-economic-research/

(The Benefits of Open Data – Evidence from Economic Research)

https://docs.google.com/presentation/d/1_uF9HSJnrS9eFgHi4gMYg1UZ_-8lfGpeutfSBd2ppKs/edit?pli=1#slide=id.p
(OKF Live Document on Evidence & Anecdotes for Open Gov Data)

http://www.nationalarchives.gov.uk/documents/nif-and-open-data.pdf

(A National Information Framework for Public Sector Information and Open Data)

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/198752/13-744-shakespeare-review-of-public-sector-information.pdf

(Shakespeare Review: An Independent Review of Public Sector Information)

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/198905/bis-13-743-market-assessment-of-public-sector-information.pdf

(Market Assessment of Public Sector Information by Deloitte)

http://vimeo.com/29259763

(Open Government Promo Video by OGP)

http://www.rooter.es/userfiles/Rooter_innovacion_en_servicios.pdf

(La Innovación en Servicios en España)

http://rooter.es/documents/PAPER_REUTILIZACION_INFORMACION_PUBLICA_PRIVADA_OPENDATA.pdf

(Reutilización de información pública y privada en España)

http://data.worldbank.org/sites/default/files/1/value-assessment-danish-address-data-uk-2010-07-07b.pdf

(The value of Danish address data: Social benefits from the 2002 agreement on procuring address data etc. free of charge)

http://data.worldbank.org/sites/default/files/1/denmark-address-epsiplatform2012-mortenlindmbbl-dkepsi2012effectofloweringpsichargesv02a1-120330052641-phpapp01.pptx

(Presentation)

http://sunlightfoundation.com/blog/tag/why-open-data/

(Sunlight Foundation: Why Open Data?)

http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information?cid=other-eml-alt-mgi-mck-oth-2910/

(The McKinsey Global Institute: Open data: Unlocking innovation and performance with liquid information)

This untapped potential can be unleashed if we turn public government data into open data. This will only happen, however, if it is really open (i.e. if there are no legal, financial, or technological restrictions) for re-use by others. Every restriction will exclude people from re-using the public data, and make it harder to find valuable ways of doing that.

> **For the potential to be realized, public data needs to be open data.**

1.3 What is Open Data?

This handbook is about open data but what exactly is it? In particular what makes open data open, and what sorts of data are we talking about?

1.3.1 What is Open?

This handbook is about open data – but what exactly is open data? For our purposes, open data is as defined by the Open Definition:

Terminology

The term **knowledge** is taken to include:

- Content such as music, films, books
- Data be it scientific, historical, geographic or otherwise
- Government and other administrative information

Software is excluded despite its obvious centrality because it is already adequately addressed by previous work.

The term **work** will be used to denote the item or piece of knowledge which is being transferred.

The term **package** may also be used to denote a collection of works. Of course such a package may be considered a work in itself.

The term **license** refers to the legal license under which the work is made available. Where no license has been made this should be interpreted as referring to the resulting default legal conditions under which the work is available (e.g., copyright).

The Definition

A work is open if its manner of distribution satisfies the following conditions:

1. Access

The work shall be available as a whole and at no more than a reasonable reproduction cost, preferably downloading via the Internet without charge. The work must also be available in a convenient and modifiable form.

Comment: This can be summarized as 'social' openness – not only are you allowed to obtain the work but you can obtain it. 'As a whole' prevents the limitation of access by indirect means, for example by only allowing access to a few items of a database at a time (material should be available in bulk as necessary). Convenient and modifiable means that material should be machine readable (rather than, for example, just human readable).

2. Redistribution

The license shall not restrict any party from selling or giving away the work either on its own or as part of a package made from works from many different sources. The license shall not require a royalty or other fee for such sale or distribution.

3. Reuse

The license must allow for modifications and derivative works and must allow them to be distributed under the terms of the original work.

Comment: Note that this clause does not prevent the use of 'viral' or share-alike licenses that require redistribution of modifications under the same terms as the original.

4. Absence of Technological Restriction

The work must be provided in such a form that there are no technological obstacles to the performance of the above activities. This can be achieved by the provision of the work in an open data format (i.e. one whose specification is publicly and freely available and which places no restrictions monetary or otherwise upon its use).

5. Attribution

The license may require as a condition for redistribution and re-use the attribution of the contributors and creators to the work. If this condition is imposed it must not be onerous. For example if attribution is required a list of those requiring attribution should accompany the work.

6. Integrity

The license may require as a condition for the work being distributed in modified form that the resulting work carry a different name or version number from the original work.

7. No Discrimination Against Persons or Groups

The license must not discriminate against any person or group of persons.

Comment: In order to get the maximum benefit from the process, the maximum diversity of persons and groups should be equally eligible to contribute to open knowledge. Therefore we forbid any open-knowledge license from locking anybody out of the process.

Comment: this is taken directly from item 5 of the OSD.

8. No Discrimination Against Fields of Endeavor

The license must not restrict anyone from making use of the work in a specific field of endeavor. For example, it may not restrict the work from being used in a business, or from being used for genetic research.

Comment: The major intention of this clause is to prohibit license traps that prevent open material from being used commercially. We want commercial users to join our community, not feel excluded from it.

Comment: this is taken directly from item 6 of the OSD.

9. Distribution of License

The rights attached to the work must apply to all to whom it is redistributed without the need for execution of an additional license by those parties.

Comment: This clause is intended to forbid closing up knowledge by indirect means such as requiring a non-disclosure agreement.

Comment: this is taken directly from item 7 of the OSD.

10. License Must Not Be Specific to a Package

The rights attached to the work must not depend on the work being part of a particular package. If the work is extracted from that package and used or distributed within the terms of the work's license, all parties to whom the work is redistributed should have the same rights as those that are granted in conjunction with the original package.

Comment: this is taken directly from item 8 of the OSD.

11. License Must Not Restrict the Distribution of Other Works

The license must not place restrictions on other works that are distributed along with the licensed work. For example, the license must not insist that all other works distributed on the same medium are open.

Comment: Distributors of open knowledge have the right to make their own choices. Note that 'share-alike' licenses are conformant since those provisions only apply if the whole forms a single work.

Comment: this is taken directly from item 9 of the OSD.

–   See more at: http://opendefinition.org/od/#sthash.HP3nD1Td.dpuf

Open data is data that can be freely used, re-used and redistributed by anyone—subject only, at most, to the requirement to attribute and share alike.

Data is open if it satisfies both conditions below:

- **Technically open:** available in a machine-readable standard format, which means it can be retrieved and meaningfully processed by a computer application

- **Legally open:** explicitly licensed in a way that permits commercial and non-commercial use and re-use without restrictions. Below is an example of terms of use in data licensing:

The full Open Definition above gives precise details as to what this means. To summarize the most important:

- **Availability and Access:** the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- **Re-use and Redistribution:** the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.

- **Universal Participation:** everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavor or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

If you're wondering why it is so important to be clear about what open means and why this definition is used, there's a simple answer: **interoperability.**

> **Interoperability denotes the ability of diverse systems and organizations to work together (inter-operate). In this case, it is the ability to interoperate - or intermix - different datasets.**

Interoperability is important because it allows for different components to work together. This ability to componentize and to 'plug together' components is essential to building large, complex systems. Without interoperability this becomes near impossible — as evidenced in the most famous myth of the Tower of Babel where the (in)ability to communicate (to interoperate) resulted in the complete breakdown of the tower-building effort.

We face a similar situation with regard to data. The core of a "commons" of data (or code) is that one piece of "open" material contained therein can be freely intermixed with other "open" material. This interoperability is absolutely key to realizing the main practical benefits of openness" the dramatically enhanced ability to combine different datasets together and thereby to develop more and better products and services (these benefits are discussed in more detail in the section on 'why' open data).

> **Providing a clear definition of openness ensures that when you get two open datasets from two different sources, you will be able to combine them together, and it ensures that we avoid our own tower of babel: lots of datasets but little or no ability to combine them together into the larger systems where the real value lies.**

1.3.2 What Data are You Talking About?

Readers have already seen examples of the sorts of data that are or may become open - and they will see more examples below. However, it will be useful to quickly outline what sorts of data are, or could be, open – and, equally importantly, what won't be open.

The key point is that when opening up data, the focus is on non-personal data, that is, data which does not contain information about specific individuals.

Similarly, for some kinds of government data, national security restrictions may apply.

Here is a sample of open data organized by sector:

- Budgets and Public Finance

- WB Open Budgets
- OpenSpending
- International Budget Partnership
- The International Aid Transparency Initiative (IATI)
- US IRS Tax Statistics

- Education
  - Edu Data Inventory
  - MyData Office of Educational Technology
  - CheckMySchool

- Health
  - The U.S. Department of Health & Human Services
  - Agency for Healthcare Research & Quality (AHRQ) Databases on healthcare cost & utilization in the U.S.
  - WB Health Data

- Nutrition
  - Food Security Open Data Challenge

- Agriculture
  - The USDA National Farmers Market Directory
  - UK Department of Agriculture and Rural Development

- Transport
  - OpenPlans
  - European Public Sector Information Platform: Transport

- Environment
  - Open Climate Data
  - New York City Environment Open Data

- Water
  - Global water database

- Extractive Industries
  - Extractive Industries Transparency Initiative

- Energy
  - US Department of Energy
  - Enel Open Data

- Geospatial
  - OpenStreetMap
  - Haiti Data geospatial information

- Information and Communication Technologies (ICT)
  - Australian ICT Open Datasets

1.4 How to Open up Data

This section forms the core of this handbook. It gives concrete, detailed advice on how data holders can open up data. We'll go through the basics, but also cover the pitfalls. Lastly, we will discuss the more subtle issues that can arise.

There are three key rules we recommend following when opening up data:

- **Keep it simple.** Start out small, simple, and fast. There is no requirement that every dataset must be made open right now. Starting out by opening up just one dataset, or even one part of a large dataset, is fine–of course, the more datasets you can open up the better.

Remember this is about innovation. Moving as rapidly as possible is good because it means you can build momentum and learn from experience–innovation is as much about failure as success and not every dataset will be useful.

- **Engage early and engage often.** Engage with actual and potential users and re-users of the data as early and as often as you can, be they citizens, businesses, or developers. This will ensure that the next iteration of your service is as relevant as it can be.

It is essential to bear in mind that much of the data will not reach ultimate users directly, but rather via "info-mediaries" These are the people who take the data and transform or remix it to be presented. For example, most of us don't want or need a large database of GPS coordinates, we would much prefer a map. Thus, engage with info-mediaries first. They will re-use and repurpose the material.

- **Address common fears and misunderstandings**. This is especially important if you are working with or within large institutions such as government. When opening up data you will encounter plenty of questions and fears. It is important to (a) identify the most important ones and (b) address them at as early a stage as possible.

There are four main steps in making data open, each of which will be covered in detail below. These are in very approximate order - many of the steps can be done simultaneously.

1. **Choose your dataset(s).** Choose the dataset(s) you plan to make open. Keep in mind that you can (and may need to) return to this step if you encounter problems at a later stage.

2. **Apply an open license.**

   (a) Determine what intellectual property rights exist in the data.

   (b) Apply a suitable 'open' license that licenses all of these rights and supports the definition of openness discussed in the section above on 'What Open Data'

   (c) NB: if you can't do this go back to step 1 and try a different dataset.

3. **Make the data available.** In bulk and in a useful format. You may also wish to consider alternative ways of making it available such as via an API.

4. **Make it discoverable. <u>Post on the web and perhaps organize a central catalog to list your open datasets.</u>**

1.4.1 Choose Dataset(s)

Choosing the dataset(s) you plan to make open is the first step – though remember that the whole process of opening up data is iterative and you can return to this step if you encounter problems later on.

If you already know exactly what dataset(s) you plan to open up you can move straight on to the next section. However, in many cases, especially for large institutions, choosing which datasets to focus on is a challenge. How should one proceed in this case?

Creating this list should be a quick process that identifies which datasets could be made open to start with. There will be time at later stages to check in detail whether each dataset is suitable.

There is **no requirement** to create a comprehensive list of your datasets. The main point to bear in mind is whether it is feasible to publish this data at all (whether openly or otherwise) - see *this previous section*.

Asking the community

We recommend that you ask the community in the first instance. That is the people who will be accessing and using the data, as they are likely to have a good understanding of which data could be valuable.

1.4.1 Prepare a short list of potential datasets that you would like feedback on. It is not essential that this list concurs with your expectations, the main intention is to get a feel for the demand. This could be based on other countries' *open data* catalogs.

1.4.2 Create a request for comment.

1.4.3 Publicize your request with a webpage. Make sure that it is possible to access the request through its own URL. That way, when shared via social media, the request can be easily found.

1.4.4 Provide easy ways to submit responses. Avoid requiring registration, as it reduces the number of responses.

1.4.5 Circulate the request to relevant mailing lists, forums and individuals, pointing back to the main webpage.

1.4.6 Run a consultation event. Make sure you run it at a convenient time where the average business person, data wrangler and official can attend.

1.4.7 Ask a politician to speak on your agency's behalf. Open data is very likely to be part of a wider policy of increasing access to government information.

Cost basis

How much money do agencies spend on the collection and maintenance of data that they hold? If they spend a great deal on a particular set of data, then it is highly likely that others would like to access it.

This argument may be fairly susceptible to concerns of freeriding. The question you will need to respond to is, "Why should other people get information for free that is so expensive?" The answer is that the expense is absorbed by the public sector to perform a particular function. The cost of sending that data, once it has been collected, to a third party is approximately nothing. Therefore, they should be charged nothing.

Ease of release

Sometimes, rather than deciding which data would be most valuable, it could be useful to take a look at which data is easiest to get into the public's hands. Small, easy releases can act as the catalyst for larger behavioral change within organizations.

Be careful with this approach however. It may be the case that these small releases are of so little value that nothing is built from them. If this occurs, faith in the entire project could be undermined.

Observe peers

Open data is a growing movement. There are likely to be many people in your area who understand what other areas are doing. Formulate a list on the basis of what those agencies are doing.

What Can We Learn From New York?

[The following section was taken from the New York Open Data Handbook]

Executive Order No. 95 provides a specific definition of "Publishable State Data" to guide covered State agencies. Publishing data on data.ny.gov involves a collaborative multi-step agency process (see Figure 1: Guidance Summary). In identifying Publishable State Data, agencies should include analyses from their executive and program staff, data coordinators, FOIL officers, data stewards/IT, public information officers, security and privacy officers, and legal counsel.

Covered State entities (and entities not covered by Executive Order 95) vary widely in terms of size, personnel, functions, responsibilities, mission, and data collected and maintained. As such, the identification and prioritization processes may vary across agencies and entities. These guidelines serve to provide assistance across a broad spectrum of agencies, with the stipulation that agencies look to their governing laws, rules, regulations, and policies in identifying and publishing "publishable state data."

Figure 1: Guidance Summary

Identification
- Program/Data Owners
- Legal Counsel
- Data Stewards / IT
- Agency Executive
- FOIL Officers
- Security & Privacy Officers
- Public Information officer

Assessment / Prioritization
- High Value
- High Quality
- Complete
- Trusted
- Accessible, Machine-Readable, Reusable
- Meets legal standards for publishable state data

Pre-Publication
- Quality Assurance Validation
- Format
- Metadata
- Data Dictionary
- Timeliness
- Overview documents
- Potential research questions

Data Set Identification

In creating a data catalogue, agencies should identify those datasets that are high value, high quality, complete, and in accordance with the definition of "Publishable State Data" within Executive Order 95. "High value" data, as defined within Executive Order 95, is that which can be used to increase the agency's accountability and responsiveness, improve public knowledge of the agency and its operations, further the mission of the agency, create economic opportunity, or respond to a need or demand identified after public consultation.

The questions in Figure 2, and below, are neither exhaustive nor may be applicable to all agencies, but serve to provide a framework to identify potential data for publication on data.ny.gov. For each question, agencies must assess whether the data falls within the definition of "Publishable State Data" and the disclosure considerations that follow.

Figure 2: Identifying Publishable Data Sets
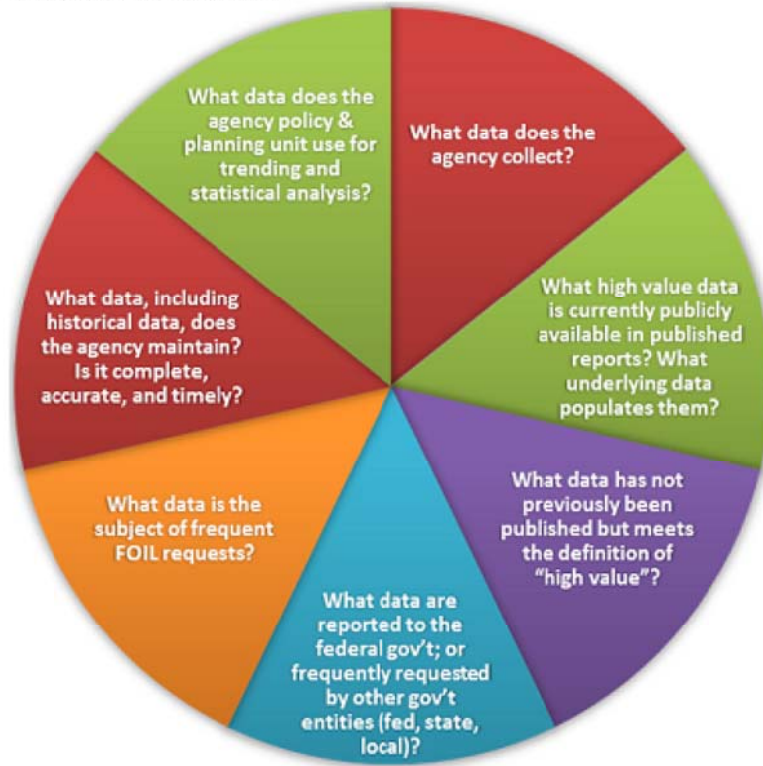
General Questions

- What data (including historical data) does the agency collect, maintain, or hold? Conducting a preliminary survey of data can be an excellent first step towards identifying publishable state data.

- What "high value" data are currently publicly available? Agencies already publish a considerable amount of data online, but it may not necessarily be accessible in bulk, or available through machine-readable mechanisms. Reviewing weekly, monthly, or quarterly reports which are frequently accessed by the public, or public-facing applications, which allow visitors to search for records, are excellent starting points.

- What underlying data populates aggregate information in published reports? Published reports are often populated with data which is compiled or aggregated from internal systems. For example, a weekly public report may indicate that an agency has closed 25 projects in that week. The internal system, which has details of each case, may have additional details which can be made public.

- What data does the agency policy and planning unit use for trending and statistical analysis? Similar to published reports, trend and statistical analysis is often performed using data from various sources. Those sources can be reviewed for data which can be made public.

- What data are reported to the federal government; or frequently requested by other government entities (federal, state, local)? Reporting of data, which is required by statute, grant, or other agreement, may already be done by an agency. Reviewing these reports (and their underlying data sources) can help identify data which can also be provided to the public. In addition, meeting these reporting requirements (particularly statutory ones) might be accomplished simply by making the data set(s) available on data.ny.gov.

- What data is the subject of frequent FOIL requests? What data is the public requesting? There are multiple methods by which the public requests data from agencies. For example, some Freedom of Information Law (FOIL) requests may seek to obtain datasets or records which are to be provided back to the requestor in digital format. These requests (particularly repeated requests for the same dataset) might be fulfilled by making the dataset(s) available on data.ny.gov.

- What data have not been previously published but meet the definition of "high value"? Publishable state data that can be used to increase the covered State entity's accountability and responsiveness, improve public knowledge of the entity and its operations, further the mission of the entity, create economic opportunity, or respond to a need or demand identified after public consultation.

Do the datasets represent discrete, usable information

In identifying datasets, government entities may be concerned that users of data.ny.gov will not understand their raw data or, if distilled to its rawest form, might lose utility. For example, state and local rules might differ, such that publishing raw, separate datasets of the two may reduce the value of the raw data being combined into a single dataset.

There are no hard and fast rules about what level of detail is sufficiently granular to add value to a government dataset.

> **Whenever possible, government entities should resist the temptation to limit datasets to only those the agency believes might be understood or useful.**

Entities should be wary of underestimating the users of data.ny.gov. data.ny.gov users may come from a variety of fields and specialties, including academic and other government users who can envision a use for the raw data not anticipated by the originating entity. A better practice is for the agency to ensure its metadata describing the dataset is complete, including comprehensive overview documents describing the data, data collection, data fields, and presentation of research questions to maximize the utility and usefulness of the data.
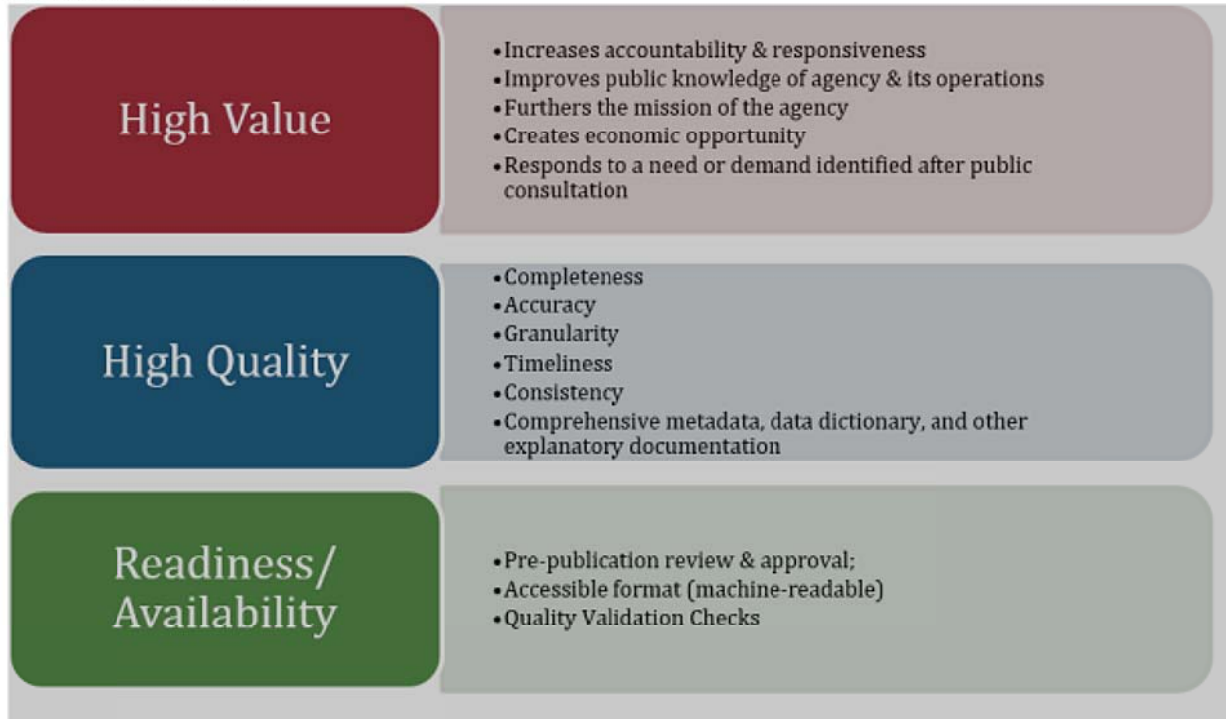
Release Prioritization

Executive Order 95 states: *"Prioritization of publication of data based on the extent to which the data can be used to increase the covered State entity's accountability and responsiveness, improve public knowledge of the entity and its operations, further the mission of the entity, create economic opportunity, or respond to a need or demand identified after public consultation…"*

Executive Order 95 further states: *"Data shall not be Publishable State Data if making such data available on the Open Data Website data.ny.gov would…impose an undue financial, operational or administrative burden on the covered State entity or State."*

When creating a schedule for publication of a particular dataset, agencies must make an assessment based upon a number of different factors. Agencies may use the guidance below to determine the priority for each data set. Prioritizing initial and ongoing publication will entail balancing high value with data quality, data availability, and data readiness. Each covered State entity shall create schedules and prioritize data publication in accordance with guidelines set forth herein, and in a timely manner, recognizing that it may take time for agencies to prepare high quality data (noting that datasets vary in complexity and, as such, can significantly vary in preparation time).

In prioritizing data for release, therefore, agencies must account for time to: identify data, assess the data (i.e., ensure consistency, timeliness, relevance, completeness, and accuracy of the data), ensure completeness of the metadata and data dictionary, review and obtain all necessary approvals to publish the data, and prepare data, metadata and requisite accompanying documentation for publication (Figure 3).

Figure 3: Prioritization

High Value	• Increases accountability & responsiveness • Improves public knowledge of agency & its operations • Furthers the mission of the agency • Creates economic opportunity • Responds to a need or demand identified after public consultation
High Quality	• Completeness • Accuracy • Granularity • Timeliness • Consistency • Comprehensive metadata, data dictionary, and other explanatory documentation
Readiness/ Availability	• Pre-publication review & approval; • Accessible format (machine-readable) • Quality Validation Checks

Below are suggested questions, the answers to which can assist agencies in prioritizing publication of high value "publishable state data" consistent with Executive Order 95:

1. Does the data highlight agency performance, or might publication of the data benefit the public by setting higher standards? The agency might be in the forefront of standards for government performance, where exposing the data might cause other agencies to raise their performance.

2. Has the data ever been published or made available in a machine-readable format so that it can be processed, analyzed, or re-used? There may exist procedures in place which can be leveraged to publish the data, such as exports for periodic department reviews, or routine exchanges of data with other agencies.

3. Is the data "high value?" While "high value" can be subjective, your agency best understands the needs of the constituency that it serves. Publishing relevant data can ultimately support those needs.

4. Does availability of the data align with new State and/or Agency initiatives? The ordering publication of any relevant datasets accordingly might be of great value.

5. Does availability of the data align with federal initiatives or exposures of federal data? There may be higher value in the agency's data if synergies can be created.

6. Can publication of the data address regulatory or grant requirements? While some data required by regulations or grants may be inappropriate for public release, publishing the data may be an acceptable way to meet those requirements and make the data accessible to the public simultaneously.

7. Does the data support decision making at the state, local, internal agency or other external agency's level, or contain information that informs public policy? Publishing such a dataset publicly can be a powerful platform

for fostering productive civic engagement and policy debate.

8. Is the data timely? What is the dataset refresh and maintenance cycle? Systems, which support the ongoing operations of an agency, are often kept up-to-date on a daily basis. Publishing raw or aggregate data drawn from these systems can provide tremendous value.

9. Does availability of the data align with legal requirements for data publication? For example, there might be statutorily-required reporting which can be satisfied by publishing datasets, without necessarily needing an extensive narrative report. If the data is collected and compiled by the agency to fulfill statutory reporting requirements, then the agency's governing laws have already determined that the data is of high value for that agency.

10. Would availability of the data improve agency-to-agency communication? Certain government functions may involve multiple agencies requiring access to similar data.

11. Could availability of the data create specific economic opportunity? In many cases, this will be unknown to the agency in advance. Some of the greatest successes of the open data movement have involved government data being commercially appropriated in useful ways, such as weather data. To the extent the agency can anticipate significant commercial use of the data, the agency may wish to order publication of such data more highly as it creates its schedule.

12. Could the data be useful for the creation of novel and useful third-party applications, mobile applications, and services? Software applications often leverage data from multiple sources to provide value to their customers. Making agency data sets available can support the delivery of greater value (and impact) through those applications.

13. Does the data further the core mission or strategic direction of the agency or multiple government entities? Publishing aggregated data (statistics, metrics, performance indicators) as well as raw data can often help an agency advance its strategic mission. In addition, data.ny.gov can serve as a conduit for efficiently sharing information with other agencies.

14. Does the data have depth and breadth of years of coverage? Release of data with high information content and quality can improve accountability and responsiveness and/or improve public knowledge of the agency and its operations.

15. Does the data have accompanying metadata and a data dictionary? Metadata and all accompanying documents should be comprehensive so as to provide a full understanding of the data and data elements to an end-user. This ensures version control, availability of contact information, and descriptive information sufficient for end-users to be able to use and interpret the data. In addition, where applicable, agencies should append disclaimers to highlight limitation of the data and/or prevent use of the data in misleading ways.

16. Is the data accurate/complete? The dataset must be sufficiently final or complete, such that it is currently publishable. Agencies should work to transform any data sets or partial data sets which are not complete or high quality so that they can eventually be published. If there is a trigger allowing the agency to publish the data at some time in the future, then scheduling publication of the data should be set accordingly.

17. Is the dataset in a format that is machine-readable or can be easily transformed? The data should be organized or formatted in a manner which is machine-readable and that can be re-used, and capable of being digitally transmitted or processed. It should be in tabular or geo-spatial form. Agencies should consider the level of effort required to transform the data to a machine-readable format and maintain it in such a format.

18. Is the data frequently requested? As demand is known and quantifiable, this should raise the value of this data for publication. If the dataset is the type that is requested through FOIL on a recurring basis, then the agency may reduce duplication and obtain efficiencies by posting data on data.ny.gov.

19. Is the data needed by the public after-hours? As demand may be known and quantifiable. Generally when there is this type of demand for the data, such datasets should be ranked, where applicable, of higher value.

20. Does the data have a direct impact on the public? The data is likely of higher value if it is already apparent there is a deep impact and interest by the public (e.g., hospital infection rates, food establishment inspection results, etc.).

21. Is the data in strong demand from constituencies? The data might be of higher value to specific, narrow interest groups which may be the agency's core constituency for those issues.

22. Is the data of timely interest? Announcements of progress or success – or reactions to public criticism - can be strongly supported by publishing related data, should it exist.

1.4.2 Apply an Open License (Legal Openness)

[The following section was taken from the Sunlight Foundation website.]

What Data Should Be Public?

1. Set the default to open

In most jurisdictions there are intellectual property rights in data that prevent third-parties from using, reusing and redistributing data without explicit permission. Even in places where the existence of rights is uncertain, it is important to apply a license simply for the sake of clarity. Thus, **if you are planning to make your data available you should put a license on it** – and if you want your data to be open this is even more important.

What licenses can you use? We recommend that for 'open' data you use one of the licenses conformant with the Open Definition and marked as suitable for data. This list (along with instructions for usage) can be found at:

• http://opendefinition.org/licenses/

A short 1-page instruction guide to applying an open data license can be found on the Open Data Commons site:

• http://opendatacommons.org/guide/

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. Preexisting 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.

1Austin, Texas Open Data Resolution, http://www.ci.austin.tx.us/edims/document.cfm?id=161941

> **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].

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. §§ 19.3119.37) 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&GU](#) [ID=2EC21911798D4499-BFAF96BDACBCD8C7&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 "highvalue" 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 approach, listing specific data sets (i.e. registered vacant properties and crime incidents) and timelines for release.

2Madison, Wisconsin Open Data Policy, [http://madison.legistar.com/ViewReport.ashx?M=R&N=Text&GID=205&ID=1201083&GUID=2EC21911798D4499BFAF96BDACB](#) [CD8C7&Title=Legislation+Text](#)

3Utah Open Data Policy, [http://le.utah.gov/~2013/bills/sbillenr/SB0283.pdf](#)

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 quasigovernmental 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 technologyrelated procurements in its 2012 open data policy[5] .

Similarly, these public protections should generally apply to quasigovernmental agencies and other similar actors, such as multistate agencies, governmentsponsored entities, publiclyfunded universities, and selfregulatory organizations (like FINRA[6]).

SAMPLE LANGUAGE:

Chicago, Illinois has a specific provision about "TechnologyRelated Procurements."

The chief data officer shall work with the chief procurement officer to develop contract provisions to promote open data policies in technologyrelated 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

Open data policy should be complementary to preexisting legislation and directives about access to public information (see Provision 2 for more details), which means taking into consideration preexisting 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

---

4Washington, DC Open Data Policy, http://www.scribd.com/fullscreen/26442622?access_key=key-20rfsh26eu0ob66xlbmu

5Chicago Open Data Executive Order, http://www.cityofchicago.org/city/en/narr/foia/open_data_executiveorder.html

6FINRA: Open data not on the table, Nancy Watzman, Sunlight Foundation, http://reporting.sunlightfoundation.com/2012/Finra_open_data/

7Washington, DC Open Data Policy, http://www.scribd.com/fullscreen/26442622?access_key=key-20rfsh26eu0ob66xlbmu

also requires agencies to specify how the information can be aggregated, generalized, or otherwise deidentified 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 balancetested 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 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:

---

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

11 The CFPB's source code policy: open and shared, Matthew Burton, http://www.consumerfinance.gov/blog/the-cfpbssourcecodepolicyopenandshared/

> **open formats "tend to promote a wide range of uses, backward and forward compatibility, and an independence from shortterm commercial interests."** - Josh Tauberer[12]

In other words, these formats are machinereadable (structured), serve searchable, sortable data, and tend to be nonproprietary 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, 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.**

12Open Data is Civic Capital: Best Practices for "Open Government Data," Josh Tauberer, http://razor.occams.info/pubdocs/opendataciviccapital.html#bestprac

13Open States Project, http://openstates.org/

14Open States Project Best Practices Wiki, https://code.google.com/p/openstates/wiki/StateBestPractices

15Open Data Handbook File Formats Overview, http://opendatahandbook.org/en/appendices/fileformats.html

16Open Data is Civic Capital: Why Data Formats Matter, Josh Tauberer, http://razor.occams.info/pubdocs/opendataciviccapital.html#format

178 Open Government Data Principles, Public.Resource.Org, https://public.resource.org/8_principles.html

18Ten Principles for Opening Up Government Information, Sunlight Foundation, http://sunlightfoundation.com/policy/documents/tenopendataprinciples/

19The Power of Information, Ed Mayo and Tom Steinberg, http://www.opsi.gov.uk/advice/poi/powerofinformation-review.pdf

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, toplevel definition of open formats could include the following two provisions.

[Data shall] be published in a nonproprietary, searchable, sortable, platformindependent, machinereadable 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 commonsense 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:

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]

20Portland, Oregon Open Data Policy, http://www.portlandonline.com/shared/cfm/image.cfm?id=275696

21The Public Online Information Act, Sunlight Foundation, http://sunlightfoundation.com/policy/poia/

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, nondiscriminatory, 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 licenserelated 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 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 licensefree, 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

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

23Utah Open Data Policy, http://le.utah.gov/~2013/bills/sbillenr/SB0283.pdf

24New Hampshire Open Data Policy, http://www.nhliberty.org/bills/view/2012/HB418

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 realworld 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 nonproprietary 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.

1.4.2 Make Data Available (Technical Openness)

Open data needs to be technically open as well as legally open. Specifically, the data needs to be available in bulk in a *machine-readable* format.

Available. Data should be priced at no more than a reasonable cost of reproduction, preferably as a free download from the Internet. This pricing model is achieved because your agency should not undertake any cost when it provides data for use.

25UK Open Data White Paper,
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/78946/CM8353_acc.pdf

26UK Open Data White Paper,
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/78946/CM8353_acc.pdf

27Madison, Wisconsin Open Data Policy,
http://madison.legistar.com/ViewReport.ashx?M=R&N=Text&GID=205&ID=1201083&GUID=2EC21911798 D4499BFAF-96BDACBCD8C7&Title=Legislation+Text

28A State by State Look at Lobbyist Disclosure, Eric Dunn, Sunlight Foundation,
http://sunlightfoundation.com/blog/2011/07/19/astatebystatelookatlobbyistdisclosure/

29Six Degrees of Corporations, Sunlight Foundation, http://sunlightfoundation.com/sixdegrees/resources/

30Vancouver Open Data Motion, http://former.vancouver.ca/ctyclerk/cclerk//20090519/documents/motionb2.pdf

In bulk. The data should be available as a complete set. If you have a register which is collected under statute, the entire register should be available for download. A web API or similar service may also be very useful, but they are not a substitutes for bulk access.

In an open, machine-readable format. Re-use of data held by the public sector should not be subject to patent restrictions. More importantly, making sure that you are providing machine-readable formats allows for greatest re-use. To illustrate this, consider statistics published as PDF (Portable Document Format) documents, often used for high quality printing. While these statistics can be read by humans, they are very hard for a computer to use. This greatly limits the ability for others to re-use that data.

Here are a few policies that will be of great benefit:

Keep it simple.
Move fast.
Be pragmatic.

In particular it is better to give out raw data now than perfect data in six months' time.

There are many different ways to make data available to others. The most natural in the Internet age is online publication. There are many variations to this model. At its most basic, agencies make their data available via their websites and a central catalog directs visitors to the appropriate source. However, there are alternatives.

When _connectivity_ is limited or the size of the data extremely large, distribution via other formats can be warranted. This section will also discuss alternatives, which can act to keep prices very low.

Online methods

Via your existing website

The system which will be most familiar to your web content team is to provide files for download from webpages. Just as you currently provide access to discussion documents, data files are perfectly happy to be made available this way.

One difficulty with this approach is that it is very difficult for an outsider to discover where to find updated information. This option places some burden on the people creating tools with your data.

Via 3rd party sites

Many repositories have become hubs of data in particular fields. For example, pachube.com is designed to connect people with sensors to those who wish to access data from them. Sites like Infochimps.com and Talis.com allow public sector agencies to store massive quantities of data for free.

Third party sites can be very useful. The main reason for this is that they have already pooled together a community of interested people and other sets of data. When your data is part of these platforms, a type of positive compound interest is created.

Wholesale data platforms already provide the infrastructure which can support the demand. They often provide analytics and usage information. For public sector agencies, they are generally free.

These platforms can have two costs. The first is independence. Your agency needs to be able to yield control to others. This is often politically, legally or operationally difficult. The second cost may be openness. Ensure that your data platform is agnostic about who can access it. Software developers and scientists use many operating systems, from smart phones to supercomputers. They should all be able to access the data.

Via FTP servers

A less fashionable method for providing access to files is via the File Transfer Protocol (FTP). This may be suitable if your audience is technical, such as software developers and scientists. The FTP system works in place of HTTP, but is specifically designed to support file transfers.

FTP has fallen out of favor. Rather than providing a website, looking through an FTP server is much like looking through folders on a computer. Therefore, even though it is fit for purpose, there is far less capacity for web development firms to charge for customization.

As torrents

BitTorrent is a system which has become familiar to policy makers because of its association with copyright infringement. BitTorrent uses files called torrents, which work by splitting the cost of distributing files between all of the people accessing those files. Instead of servers becoming overloaded, the supply increases with the demand increases. This is the reason that this system is so successful for sharing movies. It is a wonderfully efficient way to distribute very large volumes of data.

As an API

Data can be published via an *Application Programming Interface* (API). These interfaces have become very popular. They allow programmers to select specific portions of the data, rather than providing all of the data in bulk as a large file. APIs are typically connected to a database which is being updated in real-time. This means that making information available via an API can ensure that it is up to date.

Publishing raw data in bulk should be the primary concern of all open data initiatives. There are a number of costs to providing an API:

1.4.8   The price. They require much more development and maintenance than providing files.

1.4.9   The expectations. In order to foster a community of users behind the system, it is important to provide certainty. When things go wrong, you will be expected to incur the costs of fixing them.

Access to bulk data ensures that:

1. There is no dependency on the original provider of the data, meaning that if a restructure or budget cycle changes the situation, the data are still available.

2. Anyone else can obtain a copy and redistribute it. This reduces the cost of distribution away from the source agency and means that there is no single point of failure.

3. Others can develop their own services using the data, because they have certainty that the data will not be taken away from them.

Providing data in bulk allows others to use the data beyond its original purposes. For example, it allows it to be converted into a new format, linked with other resources, or versioned and archived in multiple places. While the latest version of the data may be made available via an API, raw data should be made available in bulk at regular intervals.

For example, the Eurostat statistical service has a bulk download facility offering over 4000 data files. It is updated twice a day, offers data in *Tab-separated values* (TSV) format, and includes documentation about the download facility as well as about the data files.

Another example is the District of Columbia Data Catalog, which allows data to be downloaded in CSV and XLS format in addition to live feeds of the data.

1.4.3 Make data discoverable

Open data is nothing without users. You need to be able to make sure that people can find the source material. This section will cover different approaches.

The most important thing is to provide a neutral space which can overcome both inter-agency politics and future budget cycles. Jurisdictional borders, whether sectorial or geographical, can make cooperation difficult. However, there are significant benefits in joining forces. The easier it is for outsiders to discover data, the faster new and useful tools will be built.

Existing tools

There are a number of tools which are live on the web that are specifically designed to make data more discover- able.

One of the most prominent is the DataHub and is a catalog and data store for datasets from around the world. The site makes it easy for individuals and organizations to publish material and for data users to find material they need.

In addition, there are dozens of specialist catalogs for different sectors and places. Many scientific communities have created a catalog system for their fields, as data are often required for publication.

For government

As it has emerged, orthodox practice is for a lead agency to create a catalog for the government's data. When establishing a catalog, try to create some structure which allows many departments to easily keep their own information current.

Resist the urge to build the software to support the catalog from scratch. There are free and open source software solutions (such as CKAN) which have been adopted by many governments already. As such, investing in another platform may not be needed.

There are a few things that most open data catalogs miss. Your program could consider the following:

- Providing an avenue to allow the private and community sectors to add their data. It may be worthwhile to think of the catalog as the region's catalog, rather than the regional government's catalog.

- Facilitating improvement of the data by allowing derivatives of datasets to be cataloged. For example, someone may geocode addresses and may wish to share those results with everybody. If you only allow single versions of datasets, these improvements remain hidden.

- Be tolerant of your data appearing elsewhere. That is, content is likely to be duplicated to communities of interest. If you have river level monitoring data available, then your data may appear in a catalog for hydrologists.

- Ensure that access is equitable. Try to avoid creating a privileged level of access for officials or tenured researchers as this will undermine community participation and engagement.

For civil society

Be willing to create a supplementary catalog for non-official data.

It is very rare for governments to associate with unofficial or non-authoritative sources. Officials have often gone to great expense to ensure that there will not be political embarrassment or other harm caused from misuse or overreliance on data.

> **Moreover, governments are unlikely to be willing to support activities that mesh their information with information from businesses.**

Governments are rightfully skeptical of profit motives. Therefore, an independent catalog for community groups, businesses and others may be warranted.

1.5 So I've Opened Up Some Data, Now What?

We've looked at how to make government information legally and technically reusable. The next step is to encourage others to make use of that data.

This section looks at additional things which can be done to promote data re-use.

1.5.1 Tell the world!

First and foremost, make sure that you promote the fact that you've embarked on a campaign to promote *open data* in your area of responsibility.

> **If you open up a bunch of datasets, it's definitely worth spending a bit of time to make sure that people know (or at least can find out) that you've done so.**

In addition to things like press releases, announcements on your website, and so on, you may consider:

- Contacting prominent organizations or individuals who work/are interested in this area
- Contacting relevant mailing lists or social networking groups
- Directly contacting prospective users who you know may be interested in this data

Understanding your audience

Like all public communication, engaging with the data community needs to be targeted. Like all stakeholder groups, the right message can be wasted if it is directed to the wrong area.

Digital communities tend to be very willing to share new information, yet they very rapidly consume it. Write as if your messages will be skimmed over, rather than critically examined in-depth.

Members of the tech community are less likely than the general public to use MS Windows. This means that you should not save documents in MS Office formats which can be read offline. There are two reasons for this:

- The first is that those documents will be less accessible. Rather than the document you see on your screen, readers may see an imperfect copy from an alternative.

Post your material on third-party sites

Many blogs have created a large readership in specialized topic areas. It may be worthwhile adding an article about your initiative on their site. These can be mutually beneficial. You receive more interest and they receive a free blog post in their topic area.

Making your communications more social-media friendly

It's unrealistic to expect that officials should spend long periods of time engaging with social media. However, there are several things that you can do to make sure that your content can be easily shared between technical users. Some tips:

Provide unique pages for each piece of content. When a message is shared with others, the recipient of the referral will be looking for the relevant content quickly.

Avoid making people download your press releases. Press releases are fine. They are concise messages about a particular point. However, if you require people to download the content and for it to open outside of a web browser, then fewer people will read it. Search engines are less likely to index the content. People are less likely to click to download.

Consider using an Open license for your content. Apart from providing certainty to people who wish to share your content that this is permissible, you send a message that your agency understands openness. This is bound to leave an impression far more significant to proponents of open data than any specific sentence in your press release.

Social media

It's inefficient for cash-strapped agencies to spend hours on social media sites. The most significant way that your voice can be heard through these fora is by making sure that blog posts are easily shareable. That means, before reading the next section, make sure that you have read the last. With that in mind, here are a few suggestions:

Discussion for a. Twitter has emerged as the platform of choice for disseminating information rapidly.

> **Anything tagged with #opendata will be immediately seen by thousands. LinkedIn has a large selection of groups which are targeted towards open data.**

While Facebook is excellent for a general audience, it has not received a great deal of attention in the open data community.

Link aggregators. Submit your content to the equivalent of newswires for geeks.

> **Reddit and Hacker News are the two biggest in this arena at the moment.**

To a lesser extent, Slashdot and Digg are also useful tools in this area.

These sites have a tendency to drive significant traffic to interesting material. They are also heavily focused on topic areas.

1.5.2 Getting folks in a room: Unconferences, Meetups and Barcamps

Face-to-face events can be a very effective way to encourage others to use your data. Reasons that you may consider putting on an event include:

- Finding out more about prospective re-users

- Finding out more about demand for different datasets

- Finding out more about how people want to re-use your data

- Enabling prospective re-users to find out more about what data you have

- Enabling prospective users to meet each other (e.g. so they can collaborate)

- Exposing your data to a wider audience (e.g. from blog posts or media coverage that the event may help to generate)

There are also lots of different ways of running events, and different types of events, depending on what aim you want to achieve. As well as more traditional conference models, which will include things like pre-prepared formal talks, presentations and demonstrations, there are also various kinds of participant driven events, where those who turn up may:

- Guide or define the agenda for the event

- Introduce themselves, talk about what they're interested in and what they're working on, on an ad hoc basis

- Give impromptu micro-short presentations on something they are working on

- Lead sessions on something they are interested in

There is plenty of documentation online about how to run these kinds of events, which you can find by searching for things like: unconference, barcamp, meetup, speedgeek, lightning talk, and so on. You may also find it worthwhile to contact people who have run these kinds of events in other countries, who will most likely be keen to help you out and to advise you on your event. It may be valuable to partner with another organization (e.g. a civic society organization, a news organization or an educational institution) to broaden your base participants and to increase your exposure.

1.5.3 Making things! Hackdays, prizes and prototypes

The structure of these competitions is that a number of datasets are released and programmers then have a short time-frame — running from as little as 48 hours to a few weeks—to develop applications using the data. A prize is then awarded to the best application. Competitions have been held in a number of countries including the UK, the US, Norway, Australia, Spain, Denmark, and Finland.

Examples for Competitions

Show us a better way was the first such competition in the world. It was initiated by the UK Government's "The Power of Information Taskforce" headed by Cabinet Office Minister Tom Watson in March 2008. This competition asked "What would you create with public information?" and was open to programmers from around the world, with a tempting £80,000 prize for the five best applications.

Apps for Democracy, one of the first competitions in the United States, was launched in October 2008 by Vivek Kundra, at the time Chief Technology Officer (CTO) of the District of Columbia (DC) Government. Kundra had developed the groundbreaking DC data catalog, http://data.octo.dc.gov/, which included datasets such as real-time crime feeds, school test scores, and poverty indicators. It was at the time the most comprehensive local data catalog in the world. The challenge was to make it useful for citizens, visitors, businesses and government agencies of Washington, DC.

The creative solution was to create the Apps for Democracy contest. The strategy was to ask people to build applications using the data from the freshly launched data catalog. It included an online submission for applications, many small prizes rather than a few large ones, and several different categories as well as a "People's Choice" prize. The competition was open for 30 days and cost the DC government $50,000. In return, a total of 47 iPhone, Facebook and web applications were developed with an estimated value in excess of $2,600,000 for the local economy.

The Abre Datos (Open Data) Challenge 2010. Held in Spain in April 2010, this contest invited developers to create open source applications making use of public data in just 48 hours. The competition had 29 teams of participants who developed applications that included a mobile phone program for accessing traffic information in the Basque Country, and for accessing data on buses and bus stops in Madrid, which won the first and second prizes of €3,000 and €2,000 respectively.

Nettskap 2.0. In April 2010 the Norwegian Ministry for Government Administration held "Nettskap 2.0". Norwegian developers – companies, public agencies or individuals – were challenged to come up with web-based project ideas in the areas of service development, efficient work processes, and increased democratic participation. The use of government data was explicitly encouraged. Though the application deadline was just a month later, on May 9, the Minister Rigmor Aasrud said the response was "overwhelming". In total 137 applications were received, no less than 90 of which built on the re-use of government data. A total amount of NOK 2.5 million was distributed among the 17 winners; while the total amount applied for by the 137 applications was NOK 28.4 million.

Mashup Australia. The Australian Government 2.0 Taskforce invited citizens to show why open access to Australian government information would be positive for the country's economy and social development. The contest ran from October 7th to November 13th 2009. The Taskforce released some datasets under an open license and in a range of reusable formats. The 82 applications that were entered into the contest are further evidence of the new and innovative applications which can result from releasing government data on open terms.

Conferences, Barcamps, Hackdays

One of the more effective ways for Civil Society Organizations (CSOs) to demonstrate to governments the value of opening up their datasets is to show the multiple ways in which the information can be managed to achieve social and economic benefit. CSOs that promote re-use have been instrumental in countries where there have been advances in policy and law to ensure that datasets are both technically and legally open.

The typical activities which are undertaken as part of these initiatives normally include competitions, *open government data* conferences, unconferences, workshops and hack days. These activities are often organized by the user community with data that has already been published proactively or obtained using access to information requests. In other cases, civil society advocates have worked with progressive public officials to secure new release of datasets that can be used by programmers to create innovative applications.

1.6 Socrata's 6-Phase Approach

Think About a Pilot to Start

Pilots are a great way to get your feet wet with open data. A great number of Socrata customers started out with a limited-scope, limited-time pilot, usually for three to four months, but it can last up to a year. This approach has several advantages, such as:

- Gives you an opportunity to test the technology, try out several different approaches, discover what's possible using real-world data, and socialize your new platform with your organization's leadership, your co-workers, and citizens.

- **Three to four months is enough time to run a developer event from start to finish. Having a platform on which you can support a successful hackathon, or a longer-running developer challenge with real data, user-friendly visualizations, social collaboration, and APIs, is a powerful way to put your initiative on the fast track.**

- When you launch a pilot site in beta status, you can be more nimble, set up the right expectations internally and in the community, and give your collaborators the freedom to experiment with new ideas. This can be liberating experience since a quest for perfection can impede rapid progress.

Phase 1 – Start Small

What's the first data set you should publish? Where do you start?

What if we told you that several open data initiatives started with one data set? That's right, just one. Kansas City, Missouri and the United Nations Development Program started this way. They focused on a high-value data set that they wanted to make available to their customers through multiple interfaces: downloads, APIs, visualizations, filters, embeds, and sometimes an app, all on day one.

Our point here is it does not matter where you start, as long as you start. Think like a lean startup and adopt a "Ship it!" mentality. Your community will support any meaningful first step towards transparency and information they can use.

Steps and Recommendations:

- Go for the easiest tasks first. What data is already in the public domain, on your .gov site(s), as downloadable spreadsheets, PDFs, and ZIP files? Get all that data on your new open data site and bring it to life with easy-to-create maps, charts, guided browsing, and other interactive experiences.

- Grab all the maps that are currently available on your site(s) as shapefiles and KML files and convert those quickly to online maps that are interactive.

- Get your main team of data publishers involved from the outset. Your initiative will succeed and scale a lot more quickly if you don't have publishing bottlenecks. You can empower the data stewards from the various agencies to take a leadership role within their department.

- Open up the feedback channels on day one. Get your community to help identify what's important to them. Start an organic process that will, hopefully, take on a life of its own.
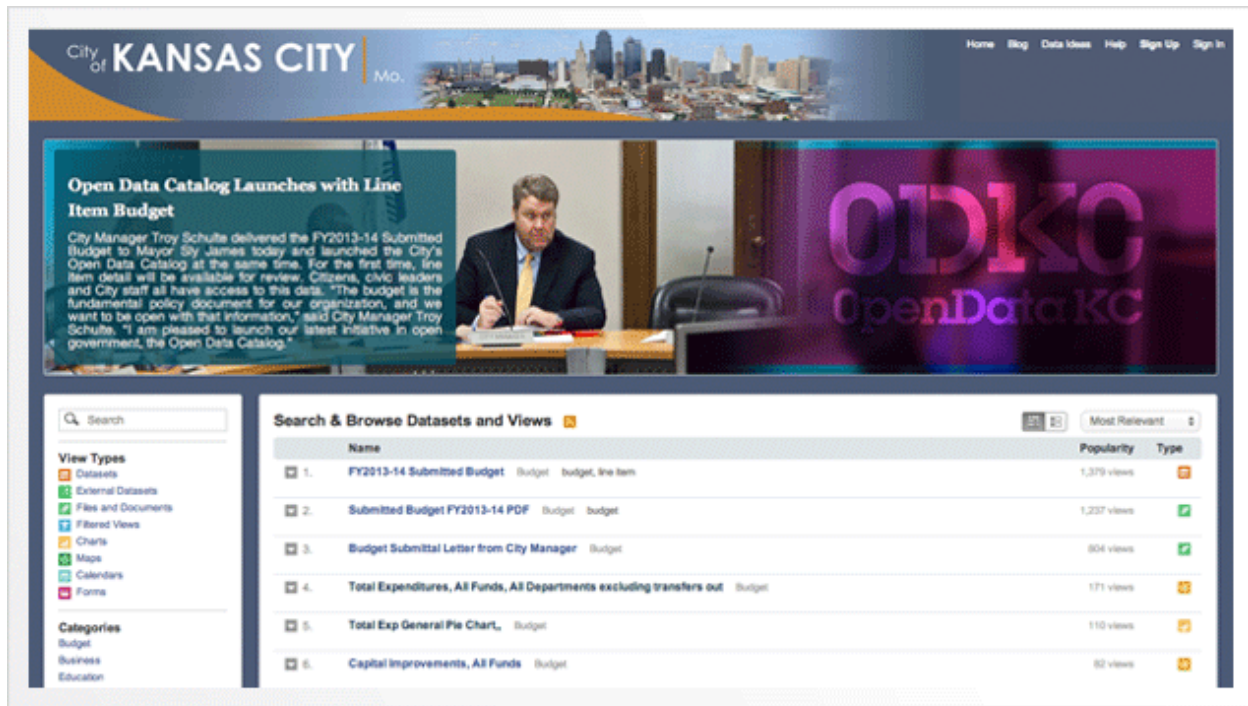
- Never worry about converting files into multiple formats or creating APIs for these data sets. That should already be taken care of for you by your open data platform.

By the end of phase one, you will hopefully have made some strides towards making public data more discoverable and usable for your constituents, and with little effort. Phase one, check!

Phase 2 – Get Transparency Done

Reaching full transparency might take a while but is essential for true open data success. Here in phase two you'll learn how to get all of the data we discussed in Chapter 6 on "The Data Plan" online, in a usable format.

Steps and Recommendations:

- **The easiest data to publish first is usually financial transparency and personnel data.** This is because it is typically well-organized and can be easily extracted from your back-end financial systems as clean spreadsheets.

- Once you load it on the platform, do a usability pass to make sure that the data makes sense to your constituents who don't know your internal codes and acronyms.

- **Build helpful visualizations, especially charts, and guided browsing filters to help people interact with and understand the data with little effort.** For example, if it is budget data, a pie chart of budget by department might be helpful. If it's expenditure data, think about building facets to browse the data by department, by month, and by type.

- You can then move on to ethics data, like campaign donations and lobbyists' salaries, which may require more clean up and data collection for your team. It's a worthwhile effort.
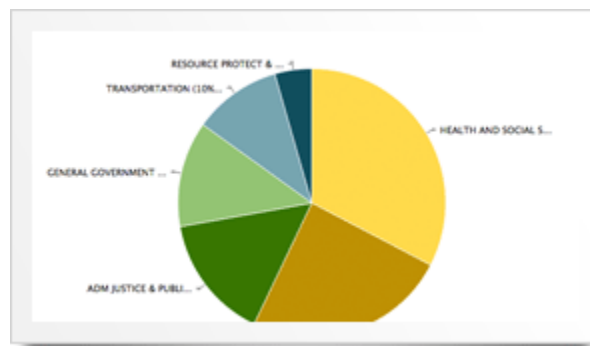
A pie chart of budget data visualized with "Open Checkbook Explorer"

Socrata's "Open Checkbook Explorer" application makes it easy to share near real-time finance data.

With [Ethics.gov](#), President Obama made good on a campaign promise to publish campaign funding and other data.

"[Chicago Lobbyists](#)" is an app developed by [Open City](#), a group of volunteer civic hackers in Chicago.

Phase 3 – Bring Developers Onboard

One of the primary goals of open data programs is to connect with your local developer community and get them what they need to build apps and other features with your data.

It's best if you take a backseat role.

> **Find open data enthusiasts in your tech community and let them lead the effort to build apps.**

Tom Schenk, director of analytics and performance management for the city of Chicago recommends "humility" in order to create an engaged developer community. He suggests letting these intelligent, creative people tell you what data would interest them most or how the data could be served up in a more useful manner.

"Chicago has the best developer community. I know many of them by their first name. They know a number of people here, from Brett Goldstein the CDO, to myself, to other developers. We talk with them on a very regular basis. Every Tuesday, we're at their meet-ups. We don't drive it, we're just there in attendance. We work on projects like everybody else."

Tom Schenk, Director of Analytics and Performance Management, City of Chicago

Steps and Recommendations:

- Familiarize yourself with successful developer and apps programs out there, such as the NYC Big Apps competition, Chicago Digital, Evergreen Apps, and ongoing developer challenges in the federal government.

- Look for local leaders in the tech community who are interested in open data. Meet with them personally. You may find them at your local colleges, or even high schools. Or, you could meet them at tech community meetings for any number of topics of interest to developers.

- If there isn't one already, start a regular meet-up for developers to discuss and work on open data projects. Choose an appealing spot with room for them to interact or do some work alone. And, try to pick a location with plenty of bandwidth so that they can all be online at the same time.

- Ask them how you can help. Maybe your 911 call data comes in a format that is difficult to work with for certain projects. Or, perhaps you're updating some data too infrequently to warrant a successful app for citizens, such as snowplow locations.

- Create opportunities for recognition, be it an annual or bi-annual award, or prizes in a hackathon. Regularly include their achievements in your press outreach about the program.

Over 40 cities, counties, and states are coordinating the first ever National Day of Civic Hacking

The state of Washington, King County, and the city of Seattle host the [Evergreen Apps Challenge](#) to "help people create useful experiences out of the data that government creates on a day-to-day basis."

The city of Chicago highlights all of the outstanding apps created by local civic hackers on its [Digital Chicago](#) site.

Check out the "[Winners Gallery](#)" for the latest New York City BigApps event.

Phase 4 – Increase Agency Participation

Your open data program will truly flourish when you can get broad participation from your agencies and departments. Getting their active participation often rests on your ability to articulate how the open data platform can help them do their jobs more efficiently and create better outcomes. Customize your message to them.

Remind them also that they are the primary beneficiaries of the Web and mobile apps that will become part of the service and information-delivery infrastructure for their programs. In addition, the program helps to break down internal silos and facilitate information sharing across agencies.

"Our goal was to improve the transparency. But, from an operational standpoint there is the added benefit of staff being able to have better access to the data. It is an efficiency gain where we can share data and build tools that find connections in the data."

Chad Janicek, Director of the Office of Management and Budget for Weatherford, Texas

Steps and Recommendations:

- **Engage with your internal stakeholders in other agencies, operating divisions, and departments.**

- Focus on agency CIOs, department heads, public information officers, communications teams, and Web managers. Ask them about the information bottlenecks in their own processes. Ask questions like, "How do they currently create and update maps, reports, apps, and other data-driven content? Do you have to rely on IT staff for every update? Are constituents satisfied with their experience and the level of access to information?"

- Show your colleagues how they go from raw data to an interactive experience on the Web in minutes. Show them how they can create visualizations and maps and embed them on their agency sites without any technical help.

- Ask them to sponsor an app for your app contest based on their needs. Maybe this is the opportunity for them to get that mobile app they've been waiting for.

- Showcase studies from other government organizations where their peers have transformed their business processes.

- Identify easy opportunities to help them, even with publishing one high-value data set. Get them to register for online courses.

Remember, transparency is served every time they publish a data set, even if that was not their primary motivation.

Phase 5: Optimize for Efficiencies and Cost Savings

Phase five is a defining phase in your open data program. Open data is not just an opportunity to increase transparency and improve the citizen experience. It is also an opportunity to save time and money by adopting new technologies that make the flow of data and information more efficient. This phase deals with practical opportunities to realize those cost savings, in IT and in every participating department.

Steps and Recommendations:

- Move requests for information to self-service channels. Start by convening a meeting with internal stakeholders who deal with public information requests, such as city and county clerks. Create a quick inventory of common, repetitive requests for information, and identify the most logical candidates for proactive disclosure in self-service delivery mode.

- These can range from FOIA-related requests, to everyday requests for maps, business listings, licenses, and permits. How much would it save your organization if 20 percent or even 10 percent of high-touch interactions could be deflected to self-service channels?

- Boost everyone's productivity. Saving information workers' time is an area where you have almost limitless potential to innovate. The productivity gains you can start implementing right away with your open data platform fall in two areas:

  1. Data and information discovery: Your employees will have a central resource where they can easily find data from their colleagues in other departments.

  2. Data collection and publishing: Your colleagues who are creating and updating interactive information on the Web – maps, charts, reports – can now do in minutes what used to take them days and weeks to do. In many cases, when you take advantage of automated publishing from your systems of record to interfaces for citizens, time-consuming manual updates can be eliminated altogether.

Read the Three Tips to IT Savings with Open Data

The Recreation and Conservation Office at the state of Washington published its "State of Salmon in Watersheds 2102 Report" using data.wa.gov *and creating 350 interactive charts.*

The city of Seattle's near real-time "911 calls" map lets citizens follow crime activity and emergencies in the city.

Cities across the U.S. are federating their data on cities.data.gov.

The city of Chicago, Cook County, and the state of Illinois have federated their data on to a single, citizen-friendly site, MetroChicagoData.gov.

Phase 6 – Federate Data with Neighboring Cities, Counties, and States

The implementation plan so far has focused on data within your own walls and the experience of your employees and constituents. This next phase deals with two-way exchanges of data with neighboring cities, counties, states, or federal agencies, as well as multi-stakeholder collaborations to create converged data portals at a regional or national level. In the world of open data, this is called data federation, or sometimes "data catalog federation."

Benefits of Federation

- Enables the comparison of quality of life and other performance indicators across states, to provide new insights and inform decision-making in our communities. A likely initial focus will be comparisons of health data.

- Creates derivative information products based on intelligent aggregation of state-level data, e.g. a national directory of community health facilities.

- Provides citizens with greater access by enabling interactive data discovery across state catalogs, using a unified search interface and normalized topical categories, such as health, economy, public safety, environment, etc.

- Lowers the bar for developers by providing consistent APIs from one jurisdiction to another and normalizing the schemas for high-value, state-level data sets that are semantically identical, such as unified crime reports.

- Increases the reach of each participant's open data efforts by developing a consumer brand that can aggregate audiences and build awareness on social networks.

- Contributes to the development of common open data standards.

1.7 Glossary

Anonymisation The process of adapting data so that individuals cannot be identified from it.

Anonymization See *Anonymisation*.

API See *Application Programming Interface*.

Application Programming Interface A way computer programs talk to one another. Can be understood in terms of how a programmer sends instructions between programs.

AR See *Information Asset Register*.

Attribution License A license that requires that the original source of the licensed material is cited (attributed).

BitTorrent BitTorrent is a protocol for distributing the bandwith for transferring very large files between the computers which are participating in the transfer. Rather than downloading a file from a specific source, BitTorrent allows peers to download from each other.

Connectivity Connectivity relates to the ability for communities to connect to the Internet, especially the World Wide Web.

Copyright A right for the creators of creative works to restrict others' use of those works. An owner of copyright is entitled to determine how others may use that work.

DAP See *Data Access Protocol*.

Data Access Protocol A system that allows outsiders to be granted access to databases without overloading either system.

Data protection legislation Data protection legislation is not about protecting the data, but about protecting the right of citizens to live without fear that information about their private lives might become public. The law protects privacy (such as information about a person's economic status, health and political position) and other rights such as the right to freedom of movement and assembly. For example, in Finland a travel card system was used to record all instances when the card was shown to the reader machine on different public transport lines. This raised a debate from the perspective of freedom of movement and the travel card data collection was abandoned based on the data protection legislation.

Database rights A right to prevent others from extracting and reusing content from a database. Exists mainly in European jurisdictions.

EU European Union.

EU PSI Directive The *Directive on the re-use of public sector information*, 2003/98/EC. "deals with the way public sector bodies should enhance re-use of their information resources." Legislative Actions - PSI Directive

IAR See *Information Asset Register*.

Information Asset Register IARs are registers specifically set up to capture and organize meta-data about the vast quantities of information held by government departments and agencies. A comprehensive IAR includes databases, old sets of files, recent electronic files, collections of statistics, research and so forth.

The *EU PSI Directive* recognizes the importance of asset registers for prospective re-users of public information. It requires member states to provide lists, portals, or something similar. It states:

Tools that help potential re-users to find documents available for re-use and the conditions for re-use can facilitate considerably the cross-border use of public sector documents. Member States should therefore ensure that practical arrangements are in place that help re-users in their search for documents available for reuse. Assets lists, accessible preferably online, of main documents (documents that are extensively re-used or that have the potential to be extensively re-used), and portal sites that are linked to decentralized assets lists are examples of such practical arrangements.

IARs can be developed in different ways. Government departments can develop their own IARs and these can be linked to national IARs. IARs can include information which is held by public bodies but which has not yet been – and maybe will not be – proactively published. Hence they allow members of the public to identify information which exists and which can be requested.

For the public to make use of these IARs, it is important that any registers of information held should be as complete as possible in order to be able to have confidence that documents can be found. The incomplete- ness of some registers is a significant problem as it creates a degree of unreliability which may discourage some from using the registers to search for information.

It is essential that the metadata in the IARs should be comprehensive so that search engines can function effectively. In the spirit of open government data, public bodies should make their IARs available to the general public as raw data under an open license so that civic hackers can make use of the data, for example by building search engines and user interfaces.

Intellectual property rights   Monopolies granted to individuals for intellectual creations.

IP rights   See *Intellectual property rights*.

Machine-readable Formats that are machine readable are ones which are able to have their data extracted by computer programs easily. PDF documents are not machine readable. Computers can display the text nicely, but have great difficulty understanding the context that surrounds the text.

Open Data   Open data are able to be used for any purpose. More details can be read at opendefinition.org.

Open Government Data *Open data* produced by the government. This is generally accepted to be data gathered during the course of business as usual activities which do not identify individuals or breach commercial sensitivity. Open government data is a subset of *Public Sector Information*, which is broader in scope. See http://opengovernmentdata.org for details.

Open standards   Generally understood as technical standards which are free from licensing restrictions. Can also be interpreted to mean standards which are developed in a vendor-neutral manner.

PSI   See *Public Sector Information*.

Public domain No copyright exists over the work. Does not exist in all jurisdictions. **Public Sector Information** Information collected or controlled by the public sector. **Re-use** Use of content outside of its original intention.

Share-alike License   A license that requires users of a work to provide the content under the same or similar conditions as the original.

Tab-separated values   Tab-separated values (TSV) are a very common form of text file format for sharing tabular data. The format is extremely simple and highly *machine-readable*.

Web API   An *API* that is designed to work over the Internet.

1.8 Appendices

1.8.1 File Formats

An Overview of File Formats

JSON

JSON is a simple file format that is very easy for any programming language to read. Its simplicity means that it is generally easier for computers to process than others, such as XML.

XML

XML is a widely used format for data exchange because it gives good opportunities to keep the structure in the data and the way files are built on, and allows developers to write parts of the documentation in with the data without interfering with the reading of them.

RDF

A W3C-recommended format called RDF makes it possible to represent data in a form that makes it easier to combine data from multiple sources. RDF data can be stored in XML and JSON, among other serializations. RDF encourages the use of URLs as identifiers, which provides a convenient way to directly interconnect existing _open data_ initiatives on the Web. RDF is still not widespread, but it has been a trend among Open Government initiatives, including the British and Spanish Government Linked Open Data projects. The inventor of the Web, Tim Berners-Lee, has recently proposed a five-star scheme that includes linked RDF data as a goal to be sought for open data initiatives.

Spreadsheets

Many authorities have information left in the spreadsheet, for example Microsoft Excel. This data can often be used immediately with the correct descriptions of what the different columns mean.

However, in some cases there can be macros and formulas in spreadsheets, which may be somewhat more cumbersome to handle. It is therefore advisable to document such calculations next to the spreadsheet, since it is generally more accessible for users to read.

Comma Separated Files

CSV files can be a very useful format because it is compact and thus suitable to transfer large sets of data with the same structure. However, the format is so Spartan that data are often useless without documentation since it can be almost impossible to guess the significance of the different columns. It is therefore particularly important for the comma-separated formats that documentation of the individual fields are accurate.

Furthermore it is essential that the structure of the file is respected, as a single omission of a field may disturb the reading of all remaining data in the file without any real opportunity to rectify it, because it cannot be determined how the remaining data should be interpreted.

Text Document

Classic documents in formats like Word, ODF, OOXML, or PDF may be sufficient to show certain kinds of data - for example, relatively stable mailing lists or equivalent. It may be cheap to exhibit in, as often it is the format the data is born in. The format gives no support to keep the structure consistent, which often means that it is difficult to enter data by automated means. Be sure to use templates as the basis of documents that will display data for re-use, so it is at least possible to pull information out of documents.

It can also support the further use of data to use typography markup as much as possible so that it becomes easier for a machine to distinguish headings (any type specified) from the content and so on. Generally it is recommended not to exhibit in word processing format, if data exists in a different format.

Plain Text

Plain text documents (.txt) are very easy for computers to read. They generally exclude structural metadata from inside the document however, meaning that developers will need to create a parser that can interpret each document as it appears.

Some problems can be caused by switching plain text files between operating systems. MS Windows, Mac OS X and other Unix variants have their own way of telling the computer that they have reached the end of the line.

Scanned image

Probably the least suitable form for most data, but both TIFF and JPEG-2000 can at least mark them with documentation of what is in the picture - right up to mark up an image of a document with full text content of the document. It may be relevant to their displaying data as images whose data are not born electronically - an obvious example is the old church records and other archival material - and a picture is better than nothing.

Proprietary formats

Some dedicated systems, etc. have their own data formats that they can save or export data in. It can sometimes be enough to expose data in such a format - especially if it is expected that further use would be in a similar system as that which they come from. Where further information on these proprietary formats can be found should always be indicated, for example by providing a link to the supplier's website. Generally it is recommended to display data in non-proprietary formats where feasible.

HTML

Nowadays much data is available in HTML format on various sites. This may well be sufficient if the data is very stable and limited in scope. In some cases, it could be preferable to have data in a form easier to download and manipulate, but as it is cheap and easy to refer to a page on a website, it might be a good starting point in the display of data.

Typically, it would be most appropriate to use tables in HTML documents to hold data, and then it is important that the various data fields are displayed and are given IDs which make it easy to find and manipulate data. Yahoo has developed a tool (http://developer.yahoo.com/yql/) that can extract structured information from a website, and such tools can do much more with the data if it is carefully tagged.

Open File Formats

Even if information is provided in electronic, machine-readable format, and in detail, there may be issues relating to the format of the file itself.

The formats in which information is published – in other words, the digital base in which the information is stored

- can either be "open" or "closed". An open format is one where the specifications for the software are available to anyone, free of charge, so that anyone can use these specifications in their own software without any limitations on re-use imposed by intellectual property rights.

If a file format is "closed", this may be either because the file format is proprietary and the specification is not publicly available, or because the file format is proprietary and even though the specification has been made public, re-use is limited. If information is released in a closed file format, this can cause significant obstacles to reusing the information encoded in it, forcing those who wish to use the information to buy the necessary software.

The benefit of open file formats is that they permit developers to produce multiple software packages and services using these formats. This then minimizes the obstacles to reusing the information they contain.

Using proprietary file formats for which the specification is not publicly available can create dependence on third- party software or file format license holders. In worst-case scenarios, this can mean that information can only be read using certain software packages, which can be prohibitively expensive, or which may become obsolete.

The preference from the *open government data* perspective therefore is that information be released in **open file formats which are machine-readable.**

Example: UK traffic data

Andrew Nicolson is a software developer who was involved in an (ultimately successful) campaign against the construction of a new road, the Westbury Eastern bypass, in the UK. Andrew was interested in accessing and using the road traffic data that was being used to justify the proposals. He managed to obtain some of the relevant data via freedom of information requests, but the local government provided the data in a proprietary format which can only be read using software produced by a company called Saturn, who specialize in traffic modelling and forecasting. There is no provision for a "read only" version of the software, so Andrew's group had no choice but to purchase a software license, eventually paying £500 (€600) when making use of an educational discount. The main software packages on the April 2010 price list from Saturn start at £13,000 (over €15,000), a price which is beyond the reach of most ordinary citizens.

Although no access to information law gives a right of access to information in open formats, open government data initiatives are starting to be accompanied by policy documents which stipulate that official information must be made available in open file formats. Setting the gold standard has been the Obama Administration, with the Open Government Directive issued in December 2009, which says:

To the extent practicable and subject to valid restrictions, agencies should publish information online in an open format that can be retrieved, downloaded, indexed, and searched by commonly used web search applications. An open format is one that is platform independent, machine readable, and made available to the public without restrictions that would impede the re-use of that information.

How do I use a given format?

When an authority must exhibit new data – data that has not been exhibited before – you should choose the format that provides the best balance between cost and suitability for purpose. For each format there are some things you should be aware of, and this section aims to explain them.

This section focuses only on how the cut surfaces are best arranged so that machines can access them directly. Advice and guidance about how web sites and web solutions should be designed can be found elsewhere.

Web services

For data that changes frequently, and where each pull is limited in size, it is very relevant to expose data through web services. There are several ways to create a web service, but some of the most used is SOAP and REST. Generally, SOAP over REST, REST services, but are very easy to develop and use, so it is a widely used standard.

Database

Like web services, databases provide direct access to data dynamically. Databases have the advantage that they can allow users to put together just the extraction that they are interested in.

There are some security concerns about allowing remote database extraction and database access is only useful if the structure of the database and the importance of individual tables and fields are well documented. Often, it is relatively simple and inexpensive to create web services that expose data from a database, which can be an easy way to address safety concerns.

1.8.2 What Legal (IP) Rights Are There in Data(bases)

When talking about data(bases) we first need to distinguish between the structure and the content of a database (when we use the term 'data' we shall mean the content of the database itself). Structural elements include things like the field names and a model for the data – the organization of these fields and their inter-relation.

In many jurisdictions it is likely that the structural elements of a database will be covered by copyright (it depends somewhat on the level of 'creativity' involved in creating this structure).

However, here we are particularly interested in the data. When we talk of data we need to be a bit careful because the word isn't particularly precise: data can mean a few items or even a single item (for example a single bibliographic record, a lat/long etc.) or data can mean a large collection (e.g. all the material in the database). To avoid confusion we shall reserve the term "content" to mean the individual items, and data to denote the collection.

Unlike for material such as text, music or film, the legal situation for data varies widely across countries. However, most jurisdictions **do** grant some rights in the data (as a collection).

This distinction between the content of a database and the collection is especially crucial for factual databases since no jurisdiction grants a monopoly right over the individual facts (the content), even though it may grant right(s) in them as a collection. To illustrate, consider the simple example of a database which lists the melting point of various substances. While the database as a whole might be protected by law so that one is not allow to access, re-use or redistribute it without permission, this would never prevent you from stating the fact that substance Y melts at temperature Z.

Forms of protection fall broadly into two cases:

- Copyright for compilations
- A *sui generis* right for collections of data

As we have already emphasized, there are no general rules and the situation varies by jurisdiction. Thus we proceed country by country detailing which (if any) of these approaches is used in a particular jurisdiction.

Finally, we should point out that in the absence of any legal protection, many providers of (closed) databases are able to use a simple contract combined with legal provisions prohibiting violation of access-control mechanisms to achieve results similar to a formal IP right. For example, if X is a provider of a citation database, it can achieve any set of terms of conditions it wants simply by:

1. Requiring users to login with a password
2. Only providing a user with an account and password on the condition that the user agrees to the terms and conditions

You can read more about the jurisdiction by jurisdiction situation in the Guide to Open Data Licensing.

1.8.3 Memorandum – Open Government Directive (December 8, 2009)

M10-06

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Peter R. Orszag, Director

SUBJECT: Open Government Directive

In the Memorandum on Transparency and Open Government, issued on January 21, 2009, the President instructed the Director of the Office of Management and Budget (OMB) to issue an Open Government Directive. Responding to that instruction, this memorandum is intended to direct executive departments and agencies to take specific actions to implement the principles of transparency, participation, and collaboration set forth in the President's Memorandum. This Directive was informed by recommendations from the Federal Chief Technology Officer, who solicited public comment through the White House Open Government Initiative.

The three principles of transparency, participation, and collaboration form the cornerstone of an open government. Transparency promotes accountability by providing the public with information about what the Government is doing. Participation allows members of the public to contribute ideas and expertise so that their government can make policies with the benefit of information that is widely dispersed in society. Collaboration improves the effectiveness of Government by encouraging partnerships and cooperation within the Federal Government, across levels of government, and between the Government and private institutions.

This Open Government Directive establishes deadlines for action. But because of the presumption of openness that the President has endorsed, agencies are encouraged to advance their open government initiatives well ahead of those deadlines. In addition to the steps delineated in this memorandum, Attorney General Eric Holder earlier this year issued new guidelines[1] for agencies with regard to the Freedom of Information Act (FOIA). With those guidelines, the Attorney General reinforced the principle that openness is the Federal Government's default position for FOIA issues.

This memorandum requires executive departments and agencies to take the following steps toward the goal of creating a more open government:

- **Publish Government Information Online**

   To increase accountability, promote informed participation by the public, and create economic opportunity, each agency shall take prompt steps to expand access to information by making it available online in open formats.[2] With respect to information, the presumption shall be in favor of openness (to the extent permitted by law and subject to valid privacy, confidentiality, security, or other restrictions).

   1. Agencies shall respect the presumption of openness by publishing information online (in addition to any other planned or mandated publication methods) and by preserving and maintaining electronic information, consistent with the Federal Records Act and other applicable law and policy. Timely publication of information is an essential component of transparency. Delays should not be viewed as an inevitable and insurmountable consequence of high demand.

   2. To the extent practicable and subject to valid restrictions, agencies should publish information online in an open format that can be retrieved, downloaded, indexed, and searched by commonly used web search applications. An open format is one that is platform independent, machine readable, and made available to the public without restrictions that would impede the re-use of that information.

   3. To the extent practical and subject to valid restrictions, agencies should proactively use modern technology to disseminate useful information, rather than waiting for specific requests under FOIA.

   4. Within 45 days, each agency shall identify and publish online in an open format at least three high-value data sets (see attachment section 3.a.i) and register those data sets via Data.gov. These must be data sets not

previously available online or in a downloadable format.

5. Within 60 days, each agency shall create an Open Government Webpage located at http://www.[agency].gov/open to serve as the gateway for agency activities related to the Open Government Directive and shall maintain and update that webpage in a timely fashion.

6. Each Open Government Webpage shall incorporate a mechanism for the public to:

    6.1. Give feedback on and assessment of the quality of published information;

    6.2. Provide input about which information to prioritize for publication; and

    6.3. Provide input on the agency's Open Government Plan (see 3.a.).

7. Each agency shall respond to public input received on its Open Government Webpage on a regular basis.

8. Each agency shall publish its annual Freedom of Information Act Report in an open format on its Open Government Webpage in addition to any other planned dissemination methods.

9. Each agency with a significant pending backlog of outstanding Freedom of Information requests shall take steps to reduce any such backlog by ten percent each year.

10. Each agency shall comply with guidance on implementing specific Presidential open government initiatives, such as Data.gov, eRulemaking, IT Dashboard, Recovery.gov, and USAspending.gov.

- **Improve the Quality of Government Information**

    To improve the quality of government information available to the public, senior leaders should make certain that the information conforms to OMB guidance on information quality[3] and that adequate systems and processes are in place within the agencies to promote such conformity.

    1. Within 45 days, each agency, in consultation with OMB, shall designate a high-level senior official to be accountable for the quality and objectivity[4] of, and internal controls over, the Federal spending information publicly disseminated through such public venues as USAspending.gov or other similar websites. The official shall participate in the agency's Senior Management Council, or similar governance structure, for the agency-wide internal control assessment pursuant to the Federal Managers' Financial Integrity Act.[5]

    2. Within 60 days, the Deputy Director for Management at OMB will issue, through separate guidance or as part of any planned comprehensive management guidance, a framework for the quality of Federal spending information publicly disseminated through such public venues as USAspending.gov or other similar websites. The framework shall require agencies to submit plans with details of the internal controls implemented over information quality, including system and process changes, and the integration of these controls within the agency's existing infrastructure. An assessment will later be made as to whether additional guidance on implementing OMB guidance on information quality is necessary to cover other types of government information disseminated to the public.

    3. Within 120 days, the Deputy Director for Management at OMB will issue, through separate guidance or as part of any planned comprehensive management guidance, a longer-term comprehensive strategy for Federal spending transparency, including the Federal Funding Accountability Transparency Act and the

American Reinvestment and Recovery Act. This guidance will identify the method for agencies to report quarterly on their progress toward improving their information quality.

- **Create and Institutionalize a Culture of Open Government**

   To create an unprecedented and sustained level of openness and accountability in every   agency, senior leaders should strive to incorporate the values of transparency, participation, and collaboration into the ongoing work of their agency. Achieving a more   open government will require the various professional disciplines within the Government – such as policy, legal, procurement, finance, and technology operations – to work   together to define and to develop open government solutions. Integration of various   disciplines facilitates organization-wide and lasting change in the way that Government works.

   1. Within 120 days, each agency shall develop and publish on its Open Government Webpage an Open Government Plan that will describe how it will improve transparency and integrate public participation and collaboration into its activities.   Additional details on the required content of this plan are attached. Each agency's plan shall be updated every two years.

   2. Within 60 days, the Federal Chief Information Officer and the Federal Chief Technology Officer shall create an Open Government Dashboard on www.whitehouse.gov/open. The Open Government Dashboard will make available each agency's Open Government Plan, together with aggregate statistics and visualizations designed to provide an assessment of the state of open government in the Executive Branch and progress over time toward meeting the deadlines for action outlined in this Directive.

   3. Within 45 days, the Deputy Director for Management at OMB, the Federal Chief Information Officer, and the Federal Chief Technology Officer will establish a working group that focuses on transparency, accountability, participation, and collaboration within the Federal Government. This group, with senior level representation from program and management offices throughout the Government, will serve several critical functions, including:

      3.1. Providing a forum to share best practices on innovative ideas to promote transparency, including system and process solutions for information collection, aggregation, validation, and dissemination;

      3.2. Coordinating efforts to implement existing mandates for Federal spending transparency, including the Federal Funding Accountability Transparency Act and the American Reinvestment and Recovery Act; and

      3.3. Providing a forum to share best practices on innovative ideas to promote participation and collaboration, including how to experiment with new technologies, take advantage of the expertise and insight of people both inside and outside the Federal Government, and form high-impact collaborations with researchers, the private sector, and civil society.

   4. Within 90 days, the Deputy Director for Management at OMB will issue, through separate guidance or as part of any planned comprehensive management guidance, a framework for how agencies can use challenges, prizes, and other incentive-backed strategies to find innovative or cost-effective solutions to improving open government.

- **Create an Enabling Policy Framework for Open Government**

   Emerging technologies open new forms of communication between a government and the people. It is important

that policies evolve to realize the potential of technology for open government.

1. Within 120 days, the Administrator of the Office of Information and Regulatory Affairs (OIRA), in consultation with the Federal Chief Information Officer and the Federal Chief Technology Officer, will review existing OMB policies, such as Paperwork Reduction Act guidance and privacy guidance, to identify impediments to open government and to the use of new technologies and, where necessary, issue clarifying guidance and/or propose revisions to such policies, to promote greater openness in government.

Nothing in this Directive shall be construed to supersede existing requirements for review and clearance of pre-decisional information by the Director of the Office of Management and Budget relating to legislative, budgetary, administrative, and regulatory materials. Moreover, nothing in this Directive shall be construed to suggest that the presumption of openness precludes the legitimate protection of information whose release would threaten national security, invade personal privacy, breach confidentiality, or damage other genuinely compelling interests.

If you have any questions regarding this memorandum, please direct them to opengov@omb.eop.gov or call Nicholas Fraser, Information Policy Branch, Office of Information and Regulatory Affairs, Office of Management and Budget at (202) 395-3785

Attachment - Open Government Plan

- **Formulating the Plan:** Your agency's Open Government Plan is the public roadmap that details how your agency will incorporate the principles of the President's January 21, 2009, Memorandum on Transparency and Open Government into the core mission objectives of your agency. The Plan should reflect the input of (a) senior policy, legal, and technology leadership in your agency and (b) the general public and open government experts. It should detail the specific actions that your agency will undertake and the timeline on which it will do so.

- **Publishing the Plan:** Consistent with the deadlines set forth in this Directive, the Plan should be published online on the agency's Open Government Webpage in an open format that enables the public to download, analyze, and visualize any information and data in the Plan.

- **Components of the Plan:**

    1. **Transparency:** Your agency's Open Government Plan should explain in detail how your agency will improve transparency. It should describe steps the agency will take to conduct its work more openly and publish its information online, including any proposed changes to internal management and administrative policies to improve transparency. Specifically, as part of your Plan to enhance information dissemination, your agency should describe how it is currently meeting its legal information dissemination obligations,[6] and how it plans to improve its existing information dissemination practices by providing:

        1.1. A strategic action plan for transparency that (1) inventories agency high-value information currently available for download; (2) fosters the public's use of this information to increase public knowledge and promote public scrutiny of agency services; and (3) identifies high value information not yet available and establishes a reasonable timeline for publication online in open formats with specific target dates. High-value information is information that can be used to increase agency accountability and responsiveness; improve public knowledge of the agency and its operations; further the core mission of the agency; create economic opportunity; or respond to need and demand as identified through public consultation.

        1.2. In cases where the agency provides public information maintained in electronic format, a plan for timely publication of the underlying data. This underlying data should be in an open format and

as granular as possible, consistent with statutory responsibilities and subject to valid privacy, confidentiality, security, or other restrictions. Your agency should also identify key audiences for its information and their needs, and endeavor to publish high-value information for each of those audiences in the most accessible forms and formats. In particular, information created or commissioned by the Government for educational use by teachers or students and made available online should clearly demarcate the public's right to use, modify, and distribute the information.

1.3. Details as to how your agency is complying with transparency initiative guidance such as Data.gov, eRulemaking, IT Dashboard, Recovery.gov, and USAspending.gov. Where gaps exist, the agency should detail the steps the agency is taking and the timing to meet the requirements for each initiative.

1.4. Details of proposed actions to be taken, with clear milestones, to inform the public of significant actions and business of your agency, such as through agency public meetings, briefings, press conferences on the Internet, and periodic national town hall meetings.

1.5. A link to a publicly available website that shows how your agency is meeting its existing records management requirements.[7] These requirements serve as the foundation for your agency's records management program, which includes such activities as identifying and scheduling all electronic records,[8] and ensuring the timely transfer of all permanently valuable records to the National Archives.

1.6. A link to a website that includes (1) a description of your staffing, organizational structure, and process for analyzing and responding to FOIA requests;(2) an assessment of your agency's capacity to analyze, coordinate, and respond to such requests in a timely manner, together with proposed changes, technological resources, or reforms that your agency determines are needed to strengthen your response processes; and (3) if your agency has a significant backlog, milestones that detail how your agency will reduce its pending backlog of outstanding FOIA requests by at least ten percent each year. Providing prompt responses to FOIA requests keeps the public apprised of specific informational matters they seek.

1.7. A description or link to a webpage that describes your staffing, organizational structure, and process for analyzing and responding to Congressional requests for information.

1.8. A link to a publicly available webpage where the public can learn about your agency's declassification programs, learn how to access declassified materials, and provide input about what types of information should be prioritized for declassification, as appropriate. Declassification of government information that no longer needs protection, in accordance with established procedures, is essential to the free flow of information.[9]

2. **Participation:** To create more informed and effective policies, the Federal Government should promote opportunities for the public to participate throughout the decision-making process. Your agency's Open Government Plan should explain in detail how your agency will improve participation, including steps your agency will take to revise its current practices to increase opportunities for public participation in and feedback on the agency's core mission activities. The specific details should include proposed changes to internal management and administrative policies to improve participation

2.1. The Plan should include descriptions of and links to appropriate websites where the public can engage in existing participatory processes of your agency.

2.2. The Plan should include proposals for new feedback mechanisms, including innovative tools and practices that create new and easier methods for public engagement.

3. **Collaboration:** Your agency's Open Government Plan should explain in detail how your agency will improve collaboration, including steps the agency will take to revise its current practices to further cooperation with other Federal and non-Federal governmental agencies, the public, and non-profit and private entities in fulfilling the agency's core mission activities. The specific details should include proposed changes to internal management and administrative policies to improve collaboration.

   3.1. The Plan should include proposals to use technology platforms to improve collaboration among people within and outside your agency.

   3.2. The Plan should include descriptions of and links to appropriate websites where the public can learn about existing collaboration efforts of your agency.

   3.3. The Plan should include innovative methods, such as prizes and competitions, to obtain ideas from and to increase collaboration with those in the private sector, non-profit, and academic communities.

4. **Flagship Initiative:** Each agency's Open Government Plan should describe at least one specific, new transparency, participation, or collaboration initiative that your agency is currently implementing (or that will be implemented before the next update of the Open Government Plan). That description should include:

   4.1. An overview of the initiative, how it addresses one or more of the three openness principles, and how it aims to improve agency operations;

   4.2. An explanation of how your agency engages or plans to engage the public and maintain dialogue with interested parties who could contribute innovative ideas to the initiative;

   4.3. If appropriate, identification of any partners external to your agency with whom you directly collaborate on the initiative;

   4.4. An account of how your agency plans to measure improved transparency, participation, and/or collaboration through this initiative; and

   4.5. An explanation of the steps your agency is taking to make the initiative sustainable and allow for continued improvement.

5. **Public and Agency Involvement:** Your agency's Open Government Plan should include, but not be limited to, the requirements set forth in this attachment. Extensive public and employee engagement should take place during the formation of this plan, which should lead to the incorporation of relevant and useful ideas developed in that dialogue. Public engagement should continue to be part of your agency's periodic review and modification of its plan. Your agency should respond to public feedback on a regular basis.

1.8.4 Memorandum – Open Data Policy - Managing Information as an Asset

M-13-13 — Memorandum for the Heads of Executive Departments and Agencies

From:

- **Sylvia M. Burwell**
  Director

- **Steven VanRoekel**
  Federal Chief Information Officer

- **Todd Park**
  U.S. Chief Technology Officer

- **Dominic J. Mancini**
  Deputy Administrator, Office of Information and Regulatory Affairs

Subject: Open Data Policy—Managing Information as an Asset

Information is a valuable national resource and a strategic asset to the Federal Government, its partners, and the public. In order to ensure that the Federal Government is taking full advantage of its information resources, executive departments and agencies (hereafter referred to as "agencies") must manage information as an asset throughout its life cycle to promote openness and interoperability, and properly safeguard systems and information. Managing government information as an asset will increase operational efficiencies, reduce costs, improve services, support mission needs, safeguard personal information, and increase public access to valuable government information.

Making information resources accessible, discoverable, and usable by the public can help fuel entrepreneurship, innovation, and scientific discovery – all of which improve Americans' lives and contribute significantly to job creation. For example, decades ago, the Federal Government made both weather data and the Global Positioning System (GPS) freely available to anyone. Since then, American entrepreneurs and innovators have used these resources to create navigation systems, weather newscasts and warning systems, location-based applications, precision farming tools, and much more.

Pursuant to the Executive Order of May 9, 2013, *Making Open and Machine Readable the New Default for Government Information*, this Memorandum establishes a framework to help institutionalize the principles of effective information management at each stage of the information's life cycle to promote interoperability and openness. Whether or not particular information can be made public, agencies can apply this framework to all information resources to promote efficiency and produce value.

Specifically, this Memorandum requires agencies to collect or create information in a way that supports downstream information processing and dissemination activities. This includes using machine-readable and open formats, data standards, and common core and extensible metadata for all new information creation and collection efforts. It also includes agencies ensuring information stewardship through the use of open licenses and review of information for privacy, confidentiality, security, or other restrictions to release. Additionally, it involves agencies building or modernizing information systems in a way that maximizes interoperability and information accessibility, maintains internal and external data asset inventories, enhances information safeguards, and clarifies information management responsibilities.

The Federal Government has already made significant progress in improving its management of information resources to increase interoperability and openness. The President's Memorandum on *Transparency and Open Government* [1] instructed agencies to take specific actions to implement the principles of transparency, participation, and collaboration, and the Office of Management and Budget's (OMB) *Open Government Directive* [2] required agencies to expand access to information by making it available online in open formats. OMB has also developed policies to help agencies incorporate sound information practices, including OMB Circular A-130 [3] and OMB Memorandum M-06-02. [4] In addition, the Federal Government launched Data.gov, an online platform designed to increase access to Federal datasets. The publication of thousands of data assets through Data.gov has enabled the development of numerous products and services that benefit the public.

To help build on these efforts, the President issued a Memorandum on May 23, 2012 entitled *Building a 21st Century Digital Government* [5] that charged the Federal Chief Information Officer (CIO) with developing and implementing a comprehensive government-wide strategy to deliver better digital services to the American people. The resulting *Digital Government Strategy* [6] outlined an information-centric approach to transform how the Federal Government builds and delivers digital services, and required OMB to develop guidance to increase the interoperability and openness of government information.

This Memorandum is designed to be consistent with existing requirements in the Paperwork Reduction Act, [7] the E-Government Act of 2002, [8] the Privacy Act of 1974, [9] the Federal Information Security Management Act of 2002 (FISMA), [10] the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA), [11] the Freedom of Information Act,

[12] the Information Quality Act, [13] the Federal Records Act, [14] and existing OMB and Office of Science and Technology Policy (OSTP) guidance.

If agencies have any questions regarding this Memorandum, please direct them to OMB at datause@omb.eop.gov.

This attachment provides definitions and implementation guidance for M-13-13, *Open Data Policy—Managing Information as an Asset.*

- **Definitions**

Data: For the purposes of this Memorandum, the term "data" refers to all structured information, unless otherwise noted. [15]

Dataset: For the purposes of this Memorandum, the term "dataset" refers to a collection of data presented in tabular or non-tabular form.

Fair Information Practice Principles: The term "Fair Information Practice Principles" refers to the eight widely accepted principles for identifying and mitigating privacy impacts in information systems, programs and processes, delineated in the National Strategy for Trusted Identities in Cyberspace. [16]

Government information: As defined in OMB Circular A-130, "government information" means information created, collected, processed, disseminated, or disposed of, by or for the Federal Government.

Information: As defined in OMB Circular A-130, the term "information" means any communication or representation of knowledge such as facts, data, or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms.

Information life cycle: As defined in OMB Circular A-130, the term "information life cycle" means the stages through which information passes, typically characterized as creation or collection, processing, dissemination, use, storage, and disposition.

Personally identifiable information: As defined in OMB Memorandum M-10-23, [17] "personally identifiable information" (PII) refers to information that can be used to distinguish or trace an individual's identity, either alone or when combined with other personal or identifying information that is linked or linkable to a specific individual. The definition of PII is not anchored to any single category of information or technology. Rather, it requires a case-by-case assessment of the specific risk that an individual can be identified. In performing this assessment, it is important for an agency to recognize that non-PII can become PII whenever additional information is made publicly available (in any medium and from any source) that, when combined with other available information, could be used to identify an individual.

Mosaic Effect: The Mosaic Effect occurs when the information in an individual dataset, in isolation, may not pose a risk of identifying an individual (or threatening some other important interest such as security), but when combined with other available information, could pose such risk. Before disclosing potential PII or other potentially sensitive information, agencies must consider other publicly available data – in any medium and from any source – to determine whether some combination of existing data and the data intended to be publicly released could allow for the identification of an individual or pose another security concern.

Open data: For the purposes of this Memorandum, the term "open data" refers to publicly available data structured in a way that enables the data to be fully discoverable and usable by end users. In general, open data will be consistent with the following principles:

- *Public.* Consistent with OMB's *Open Government Directive*, agencies must adopt a presumption in favor of openness to the extent permitted by law and subject to privacy, confidentiality, security, or other valid restrictions.

- *Accessible.* Open data are made available in convenient, modifiable, and open formats that can be retrieved, downloaded, indexed, and searched. Formats should be machine-readable (i.e., data are reasonably structured to allow automated processing). Open data structures do not discriminate against any person or group of persons and should be made available to the widest range of users for the widest range of purposes, often by providing the data in multiple formats for consumption. To the extent permitted by law, these formats should be non-proprietary, publicly available, and no restrictions should be placed upon their use.

- *Described.* Open data are described fully so that consumers of the data have sufficient information to understand their strengths, weaknesses, analytical limitations, security requirements, as well as how to process them. This involves the use of robust, granular metadata (i.e., fields or elements that describe data), thorough documentation of data elements, data dictionaries, and, if applicable, additional descriptions of the purpose of the collection, the population of interest, the characteristics of the sample, and the method of data collection.

- *Reusable.* Open data are made available under an open license that places no restrictions on their use.

- *Complete.* Open data are published in primary forms (i.e., as collected at the source), with the finest possible level of granularity that is practicable and permitted by law and other requirements. Derived or aggregate open data should also be published but must reference the primary data.

- *Timely.* Open data are made available as quickly as necessary to preserve the value of the data. Frequency of release should account for key audiences and downstream needs.

- *Managed Post-Release.* A point of contact must be designated to assist with data use and to respond to complaints about adherence to these open data requirements.

- *Project Open Data:* "Project Open Data," a new OMB and OSTP resource, is an online repository of tools, best practices, and schema to help agencies adopt the framework presented in this guidance. Project Open Data can be accessed at project-open-data.github.io. [18] Project Open Data will evolve over time as a community resource to facilitate adoption of open data practices. The repository includes definitions, code, checklists, case studies, and more, and enables collaboration across the Federal Government, in partnership with public developers, as applicable. Agencies can visit Project Open Data for a more comprehensive glossary of terms related to open data.

- **Scope**

The requirements in part III, sections 1 and 2 of this Memorandum apply to all new information collection, creation, and system development efforts as well as major modernization projects that update or re-design existing information systems. National Security Systems, as defined in 40 U.S.C. § 11103, are exempt from the requirements of this policy. The requirements in part III, section 3 apply to management of all datasets used in an agency's information systems. Agencies are also encouraged to improve the discoverability and usability of existing datasets by making them "open" using the methods

outlined in this Memorandum, prioritizing those that have already been released to the public or otherwise deemed high-value or high-demand through engagement with customers (see part III, section 3.c). Agencies should exercise judgment before publicly distributing data residing in an existing system by weighing the value of openness against the cost of making those data public.

- **Policy Requirements**

Agencies management of information resources must begin at the earliest stages of the planning process, well before information is collected or created. Early strategic planning will allow the Federal Government to design systems and develop processes that unlock the full value of the information, and provide a foundation from which agencies can continue to manage information throughout its life cycle.

Agencies shall take the following actions to improve the management of information resources throughout the information's life cycle and reinforce the government's presumption in favor of openness:

- Collect or create information in a way that supports downstream information processing and dissemination activities

Consistent with OMB Circular A-130, agencies must consider, at each stage of the information life cycle, the effects of decisions and actions on other stages of the life cycle. Accordingly, to the extent permitted by law, agencies must design new information collection and creation efforts so that the information collected or created supports downstream interoperability between information systems and dissemination of information to the public, as appropriate, without the need for costly retrofitting. This includes consideration and consultation of key target audiences for the information when determining format, frequency of update, and other information management decisions. Specifically, agencies must incorporate the following requirements into future information collection and creation efforts:

1. Use machine-readable and open formats [19]

Agencies must use machine-readable and open formats for information as it is collected or created. While information should be collected electronically by default, machine-readable and open formats must be used in conjunction with both electronic and telephone or paper-based information collection efforts. Additionally, in consultation with the best practices found in Project Open Data and to the extent permitted by law, agencies should prioritize the use of open formats that are non-proprietary, publicly available, and that place no restrictions upon their use.

2. Use data standards

Consistent with existing policies relating to Federal agencies' use of standards [20] for information as it is collected or created, agencies must use standards in order to promote data interoperability and openness.

3. Ensure information stewardship through the use of open licenses

Agencies must apply open licenses, in consultation with the best practices found in Project Open Data, to information as it is collected or created so that if data are made public there are no restrictions on copying, publishing, distributing, transmitting, adapting, or otherwise using the information for non-commercial or for commercial purposes. [21] When information is acquired or accessed by an agency through performance of a contract, appropriate existing clauses [22] shall be utilized to meet these objectives while recognizing that contractors may have proprietary interests in such information, and that protection of such information may be necessary to encourage qualified contractors to participate in and apply innovative concepts to government programs.

4. Use common core and extensible metadata

Agencies must describe information using common core metadata, in consultation with the best practices found in Project Open Data, as it is collected and created. Metadata should also include information about origin, linked data, geographic location, time series continuations, data quality, and other relevant indices that reveal relationships between datasets and allow the public to determine the fitness of the data source. Agencies may expand upon the basic common metadata based on standards, specifications, or formats developed within different communities (e.g., financial, health, geospatial, law enforcement). Groups that develop and promulgate these metadata specifications must review them for compliance with the common core metadata standard, specifications, and formats.

o Build information systems to support interoperability and information accessibility

Through their acquisition and technology management processes, agencies must build or modernize information systems in a way that maximizes interoperability and information accessibility, to the extent practicable and permitted by law. To this end, agencies should leverage existing Federal IT guidance, such as the *Common Approach to Federal Enterprise Architecture*, [23] when designing information systems. Agencies must exercise forethought when architecting, building, or substantially modifying an information system to facilitate public distribution, where appropriate. In addition, the agency's CIO must validate that the following minimum requirements have been incorporated into acquisition planning documents and technical design for all new information systems and those preparing for modernization, as appropriate:

3. The system design must be scalable, flexible, and facilitate extraction of data in multiple formats and for a range of uses as internal and external needs change, including potential uses not accounted for in the original design. In general, this will involve the use of standards and specifications in the system design that promote industry best practices for information sharing, and separation of data from the application layer to maximize data reuse opportunities and incorporation of future application or technology capabilities, in consultation with the best practices found in Project Open Data;

4. All data outputs associated with the system must meet the requirements described in part III, sections 1.a-e of this Memorandum and be accounted for in the data inventory described in part III section 3.a; and

5. Data schema and dictionaries have been documented and shared with internal partners and the public, as applicable.

o Strengthen data management and release practices

To ensure that agency data assets are managed and maintained throughout their life cycle, agencies must adopt effective data asset portfolio management approaches. Within six (6) months of the date of this Memorandum, agencies and inter-agency groups must review and, where appropriate, revise existing policies and procedures to strengthen their data management and release practices to ensure consistency with the requirements in this Memorandum, and take the following actions:

1. Create and maintain an enterprise data inventory

Agencies must update their inventory of agency information resources (as required by OMB Circular A-130) [24] to include an enterprise data inventory, if it does not already exist, that accounts for datasets used in the agency's information systems. The inventory will be built out over time, with the ultimate goal of including all agency datasets, to the extent practicable. The inventory will indicate, as appropriate, if the agency has determined that the individual datasets may be made publicly available (i.e., release is permitted by law, subject to all privacy, confidentiality, security, and other valid requirements) and whether they are currently available to the public. The Senior Agency

Official for Records Management should be consulted on integration with the records management process. Agencies should use the Data Reference Model from the Federal Enterprise Architecture [25] to help create and maintain their inventory. Agencies must describe datasets within the inventory using the common core and extensible metadata (see part III, section 1.e).

2. Create and maintain a public data listing

    Any datasets in the agency's enterprise data inventory that can be made publicly available must be listed at www.[agency].gov/data in a human- and machine-readable format that enables automatic aggregation by Data.gov and other services (known as "harvestable files"), to the extent practicable. This should include datasets that can be made publicly available but have not yet been released. This public data listing should also include, to the extent permitted by law and existing terms and conditions, datasets that were produced through agency-funded grants, contracts, and cooperative agreements (excluding any data submitted primarily for the purpose of contract monitoring and administration), and, where feasible, be accompanied by standard citation information, preferably in the form of a persistent identifier. The public data listing will be built out over time, with the ultimate goal of including all agency datasets that can be made publicly available. See Project Open Data for best practices, tools, and schema to implement the public data listing and harvestable files.

3. Create a process to engage with customers to help facilitate and prioritize data release

    Agencies must create a process to engage with customers, through their www.[agency].gov/data pages and other necessary means, to solicit help in prioritizing the release of datasets and determining the most usable and appropriate formats for release. [26] Agencies should make data available in multiple formats according to their customer needs. For example, high-volume datasets of interest to developers should be released using bulk downloads as well as Application Programming Interfaces (APIs). In addition, customer engagement efforts should help agencies prioritize efforts to improve the discoverability and usability of datasets that have already been released to the public but are not yet fully "open" (e.g., they are only available in closed, inaccessible formats). See Project Open Data for best practices and tools that can be used to implement customer engagement efforts.

4. Clarify roles and responsibilities for promoting efficient and effective data release practices

    Agencies must ensure that roles and responsibilities are clearly designated for the promotion of efficient and effective data release practices across the agency, and that proper authorities have been granted to execute on related responsibilities, including:

    4.1. Communicating the strategic value of open data to internal stakeholders and the public;

    4.2. Ensuring that data released to the public are open (as defined in part I), as appropriate, and a point of contact is designated to assist open data use and to respond to complaints about adherence to open data requirements;

    4.3. Engaging entrepreneurs and innovators in the private and nonprofit sectors to encourage and facilitate the use of agency data to build applications and services;

    4.4. Working with agency components to scale best practices from bureaus and offices that excel in open data practices across the enterprise;

4.5. Working with the agency's Senior Agency Official for Privacy (SAOP) or other relevant officials to ensure that privacy and confidentiality are fully protected; and

4.6. Working with the Chief Information Security Officer (CISO) and mission owners to assess overall organizational risk, based on the impact of releasing potentially sensitive data, and make a risk-based determination.

o   Strengthen measures to ensure that privacy and confidentiality are fully protected and that data are properly secured

Agencies must incorporate privacy analyses into each stage of the information's life cycle. In particular, agencies must review the information collected or created for valid restrictions to release to determine whether it can be made publicly available, consistent with the *Open Government Directive's* presumption in favor of openness, and to the extent permitted by law and subject to privacy, confidentiality pledge, security, trade secret, contractual, or other valid restrictions to release. If the agency determines that information should not be made publicly available on one of these grounds, the agency must document this determination in consultation with its Office of General Counsel or equivalent.

As agencies consider whether or not information may be disclosed, they must also account for the "mosaic effect" of data aggregation. Agencies should note that the mosaic effect demands a risk-based analysis, often utilizing statistical methods whose parameters can change over time, depending on the nature of the information, the availability of other information, and the technology in place that could facilitate the process of identification. Because of the complexity of this analysis and the scope of data involved, agencies may choose to take advantage of entities in the Executive Branch that may have relevant expertise, including the staff of Data.gov. Ultimately, it is the responsibility of each agency to perform the necessary analysis and comply with all applicable laws, regulations, and policies. In some cases, this assessment may affect the amount, type, form, and detail of data released by agencies.

As OMB has noted, "The individual's right to privacy must be protected in Federal Government information activities involving personal information." [27] As agencies consider security-related restrictions to release, they should focus on information confidentiality, integrity, and availability as part of the agency's overall risk management framework. They are required to incorporate the National Institute of Standards and Technology (NIST) Federal Information Processing Standard (FIPS) Publication 199 "Standards for Security Categorization of Federal Information and Information Systems," which includes guidance and definitions for confidentiality, integrity, and availability. [28] Agencies should also consult with the Controlled Unclassified Information (CUI) program to ensure compliance with CUI requirements, [29] the National Strategy for Information Sharing and Safeguarding, [30] and the best practices found in Project Open Data. In addition to complying with the Privacy Act of 1974, the E-Government Act of 2002, FISMA, and CIPSEA, agencies should implement information policies based upon Fair Information Practice Principles and NIST guidance on Security and Privacy Controls for Federal Information Systems and Organizations. [31] For example, agencies must:

1.   Collect or create only that information necessary for the proper performance of agency functions and which has practical utility; [32]

2.   Limit the collection or creation of information which identifies individuals to that which is legally authorized and necessary for the proper performance of agency functions; [33]

3.   Limit the sharing of information that identifies individuals or contains proprietary information to that which is legally authorized, and impose appropriate conditions on use where a continuing obligation to ensure the confidentiality of the information exists; [34]

4.   Ensure that information is protected commensurate with the risk and magnitude of the harm that would result from the loss, misuse, or unauthorized access to or modification of such information;

[35] and

5. Take into account other publicly available information when determining whether particular information should be considered PII (as defined in part I of this Memorandum).

o Incorporate new interoperability and openness requirements into core agency processes

Consistent with 44 U.S.C. 3506 (b)(2), agencies must develop and maintain an Information Resource Management (IRM) Strategic Plan. IRM Strategic Plans should align with the agency's Strategic Plan (as required by OMB Circular A-11), [36] support the attainment of agency priority goals as mandated by the Government Performance and Results Modernization Act of 2010, [37] provide a description of how IRM activities help accomplish agency missions, and ensure that IRM decisions are integrated with organizational planning, budget, procurement, financial management, human resources management, and program decisions. As part of the annual PortfolioStat process, [38] agencies must update their IRM Strategic Plans to describe how they are meeting new and existing information life cycle management requirements. Specifically, agencies must describe how they have institutionalized and operationalized the interoperability and openness requirements in this Memorandum into their core processes across all applicable agency programs and stakeholders.

• Implementation

As agencies take steps to meet the requirements in this Memorandum, it is important to work strategically and prioritize those elements that can be addressed immediately, support mission-critical objectives, and result in more efficient use of taxpayer dollars.

Agencies should consider the following as they implement the requirements of this Memorandum:

o Roles and Responsibilities

The Clinger-Cohen Act of 1996 assigns agency CIOs statutory responsibility for promoting the effective and efficient design and operation of all major IRM processes within their agency. Accordingly, agency heads must ensure that CIOs are positioned with the responsibility and authority to implement the requirements of this Memorandum in coordination with the agency's Chief Acquisition Officer, Chief Financial Officer, Chief Technology Officer, Senior Agency Official for Geospatial Information, Senior Agency Official for Privacy (SAOP), Chief Information Security Officer (CISO), Senior Agency Official for Records Management, and Chief Freedom of Information Act (FOIA) Officer. The CIO should also work with the agency's public affairs staff, open government staff, web manager or digital strategist, program owners and other leadership, as applicable.

A key component of agencies' management of information resources involves working closely with the agency's SAOP and other relevant officials to ensure that each stage of the planning process includes a full analysis of privacy, confidentiality, and security issues. Agency heads must also ensure that privacy and security officials are positioned with the authority to identify information that may require additional protection and agency activities that may require additional safeguards. Consistent with OMB Memorandum M-05-08, [39] each agency's SAOP must take on a central planning and policy-making role in all agency information management activities, beginning at the earliest stages of planning and continuing throughout the life cycle of the information. In addition, if an agency's SAOP is not positioned within the office of the CIO, the agency should designate an official within the office of the CIO to serve as a liaison to help coordinate with the agency's privacy office.

o Government-wide Coordination

The Federal CIO will work with the United States Chief Technology Officer (CTO) and the Administrator of the OMB Office of Information and Regulatory Affairs (OIRA) to help improve the interoperability and openness of government information. To this end, the Federal CIO will work to establish an inter-agency working group supported by the Federal CIO Council. The working group should focus on leveraging government-wide communities of practice to help with the development of tools that support information interoperability and openness through repositories such as Project Open Data. Part of this work should be to share best practices related to interoperability and openness within government (e.g., Federal, state, local, and tribal). These collaborations shall be subject to statutory limitations and conducted in a way that fully protects privacy, confidentiality, confidential business information, and intellectual property rights.

- o Resources

Policy implementation may require upfront investments depending on the maturity of existing information life cycle management processes at individual agencies. Agencies are encouraged to evaluate current processes and identify implementation opportunities that may result in more efficient use of taxpayer dollars. However, effective implementation should result in downstream cost savings for the enterprise through increased interoperability and accessibility of the agency's information resources. Therefore, these potential upfront investments should be considered in the context of their future benefits and be funded appropriately through the agency's capital planning and budget processes. Some of the requirements in this Memorandum may require additional tools and resources. Agencies should make progress commensurate with available tools and resources.

In addition, tools, best practices, and schema to help agencies implement the requirements of this Memorandum can be found through the Digital Services Innovation Center and in Project Open Data.

- o Accountability Mechanisms

Progress on agency implementation of the actions required in this Memorandum will be primarily assessed by OMB and the public through analysis of the agency's updates to IRM plans (part III, section 5), the completeness of the enterprise data inventory (part III, section 3.a.), and the data made available in the agency's public data listing (part III, section 3.b.).

Nothing in this Memorandum shall be construed to affect existing requirements for review and clearance of pre-decisional information by OMB relating to legislative, budgetary, administrative, and regulatory materials. Moreover, nothing in this Memorandum shall be construed to reduce the protection of information whose release would threaten national security, invade personal privacy, breach confidentiality or contractual terms, violate the Trade Secrets Act, [40] violate other statutory confidentiality requirements, [41] or damage other compelling interests. This Memorandum is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or in equity by any party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person.

1.8.5 Executive Order – Making Open and Machine Readable the New Default for Government Information

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. General Principles.

Openness in government strengthens our democracy, promotes the delivery of efficient and effective services to the public, and contributes to economic growth. As one vital benefit of open government, making information resources easy to find, accessible, and usable can fuel entrepreneurship, innovation, and scientific discovery that improves Americans' lives and contributes significantly to job creation.

Decades ago, the U.S. Government made both weather data and the Global Positioning System freely available. Since that time, American entrepreneurs and innovators have utilized these resources to create navigation systems, weather newscasts and warning systems, location-based applications, precision farming tools, and much more, improving Americans' lives in countless ways and leading to economic growth and job creation. In recent years, thousands of Government data resources across fields such as health and medicine, education, energy, public safety, global development, and finance have been posted in machine-readable form for free public use on Data.gov. Entrepreneurs and innovators have continued to develop a vast range of useful new products and businesses using these public information resources, creating good jobs in the process.

To promote continued job growth, Government efficiency, and the social good that can be gained from opening Government data to the public, the default state of new and modernized Government information resources shall be open and machine readable. Government information shall be managed as an asset throughout its life cycle to promote interoperability and openness, and, wherever possible and legally permissible, to ensure that data are released to the public in ways that make the data easy to find, accessible, and usable. In making this the new default state, executive departments and agencies (agencies) shall ensure that they safeguard individual privacy, confidentiality, and national security.

Section 2. Open Data Policy.

(a) The Director of the Office of Management and Budget (OMB), in consultation with the Chief Information Officer (CIO), Chief Technology Officer (CTO), and Administrator of the Office of Information and Regulatory Affairs (OIRA), shall issue an Open Data Policy to advance the management of Government information as an asset, consistent with my memorandum of January 21, 2009 (Transparency and Open Government), OMB Memorandum M-10-06 (Open Government Directive), OMB and National Archives and Records Administration Memorandum M-12-18 (Managing Government Records Directive), the Office of Science and Technology Policy Memorandum of February 22, 2013 (Increasing Access to the Results of Federally Funded Scientific Research), and the CIO's strategy entitled "Digital Government: Building a 21st Century Platform to Better Serve the American People." The Open Data Policy shall be updated as needed.

(b) Agencies shall implement the requirements of the Open Data Policy and shall adhere to the deadlines for specific actions specified therein. When implementing the Open Data Policy, agencies shall incorporate a full analysis of privacy, confidentiality, and security risks into each stage of the information lifecycle to identify information that should not be released. These review processes should be overseen by the senior agency official for privacy. It is vital that agencies not release information if doing so would violate any law or policy, or jeopardize privacy, confidentiality, or national security.

Section 3. Implementation of the Open Data Policy.

To facilitate effective Government-wide implementation of the Open Data Policy, I direct the following:

(a) Within 30 days of the issuance of the Open Data Policy, the CIO and CTO shall publish an open online repository of tools and best practices to assist agencies in integrating the Open Data Policy into their operations in furtherance of their missions. The CIO and CTO shall regularly update this online repository as needed to ensure it remains a

resource to facilitate the adoption of open data practices.

(b) Within 90 days of the issuance of the Open Data Policy, the Administrator for Federal Procurement Policy, Controller of the Office of Federal Financial Management, CIO, and Administrator of OIRA shall work with the Chief Acquisition Officers Council, Chief Financial Officers Council, Chief Information Officers Council, and Federal Records Council to identify and initiate implementation of measures to support the integration of the Open Data Policy requirements into Federal acquisition and grant-making processes. Such efforts may include developing sample requirements language, grant and contract language, and workforce tools for agency acquisition, grant, and information management and technology professionals.

(c) Within 90 days of the date of this order, the Chief Performance Officer (CPO) shall work with the President's Management Council to establish a Cross-Agency Priority (CAP) Goal to track implementation of the Open Data Policy. The CPO shall work with agencies to set incremental performance goals, ensuring they have metrics and milestones in place to monitor advancement toward the CAP Goal. Progress on these goals shall be analyzed and reviewed by agency leadership, pursuant to the GPRA Modernization Act of 2010 (Public Law 111-352).

(d) Within 180 days of the date of this order, agencies shall report progress on the implementation of the CAP Goal to the CPO. Thereafter, agencies shall report progress quarterly, and as appropriate.

Section 4. General Provisions.

(a) Nothing in this order shall be construed to impair or otherwise affect:

    a.i.   The authority granted by law to an executive department, agency, or the head thereof; or

    a.ii.   The functions of the Director of OMB relating to budgetary, administrative, or legislative proposals.

(b) This order shall be implemented consistent with applicable law and subject to the availability of appropriations.

(c) This order is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or in equity by any party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person.

(d) Nothing in this order shall compel or authorize the disclosure of privileged information, law enforcement information, national security information, personal information, or information the disclosure of which is prohibited by law.

(e) Independent agencies are requested to ad is here to this order.

BARACK OBAMA