

# 3

## Open Governments, Open Data

### *Moving toward a Digital Commons Framework*

*Anjanette H. Raymond and Inna Kouper*

#### INTRODUCTION

Governments around the world have embraced technology deployments intended to make cities, communities, and governments smarter. Yet the wealth of data that governments and communities gather is often left in databases or is simply released as files in an online portal. The commitment to this simplified portal-based data sharing has led some to question if governments have data management plans to assist in deciding what data should be widely open and publicly available in a highly connected, always available online world where data is easy to collect, but difficult to delete or challenge.

To examine approaches to managing government-held data in a structured and detailed way, this chapter will apply the Governing Knowledge Commons (GKC) framework in order to analyze open government data (OGD) as a knowledge resource (Frischmann, Madison, and Strandburg 2014). Specific to our case, we argue that a framework that considers data as belonging to digital commons of intellectual and cultural resources is an appropriate and useful tool to consider open government data. As Divya Siddarth and E. Glen Weyl (2021) point out:

Data lends itself especially well to a commons framework: both inputs and impacts are fundamentally shared, distributing access to these resources provides a foundation for further bottom-up innovation and technological progress, siloing or privatizing these erodes the possibility of stewarding collective benefit. Together, they form a shared layer necessary for economic growth and democratic participation.

They go on to argue for a paradigm shift in the way we think about technology and economic power:

Many core technologies increasingly resemble resources like air, water and a habitable earth – resources that are expected to be accessible to, and managed in the interest of, all members of society. In other words – a *commons-based resource*.

Taking inspiration from the Ostromian approach, shifting our view to see technology as a *digital commons* could begin to create more flexible, responsive and regenerative systems to build and deploy technology. (Siddarth and Weyl 2021)

Digital data is an essential part of the technology ecosystem. It became a ubiquitous essential resource that drives economic and technological development and, as such, it needs to be managed in the interest of all. In some instances, data likely should be considered digital public goods, non rivalrous and non excludable resources that can be created and distributed without becoming depleted for free or at a low cost (Digital Public Goods Alliance n.d.; Greenstein 2013). To reconsider data and related technologies in the hands of governments as digital public goods that belong to the commons, the conversation must address the challenges and responsibilities of collecting and managing OGD.

Many of the digital public good conceptualizations are already captured at the federal level in the United States. For example, the federal circular entitled “Managing Federal Information as a Strategic Resource” summarizes:

Federal information is both a strategic asset and a valuable national resource. It enables the Government to carry out its mission and programs effectively. It provides the public with knowledge of the Government, society, economy, and environment – past, present, and future. Federal information is also a means to ensure the accountability of Government, to manage the Government’s operations, and to maintain and enhance the performance of the economy, the public health, and welfare. Appropriate access to Federal information significantly enhances the value of the information and the return on the Nation’s investment in its creation. (Office of Management and Budget (OMB) 2016)

Simply put, data and information are valuable for many reasons in the government’s use of smart and innovative technologies. In addition to driving technological innovation, open data can increase governmental accountability and efficiency; it is one of the drivers behind the move toward improved decisions, equity, and progress. Yet much of the data held by governments remains in a troubling state. With unfunded mandates, open data portals do not receive sufficient attention beyond simple technical implementation. One might argue that government’s commitment to “open data” is largely a failure, particularly when it comes to achieving impacts beyond access to information (Attard et al. 2015).

#### OPEN DATA AND OPEN GOVERNMENT

Underlying the idea of open government there is a long-standing, well-researched tradition surrounding the right to know as a US constitutional principle (Schudson 2018). Even though the principles of open government often need to be reconciled with certain information restrictions, US administrations tend to publicly uphold and even expand their commitment to these principles. As President Obama wrote

in his Executive Order “Making Open and Machine Readable the New Default for Government Information”: “My Administration is committed to creating an unprecedented level of openness in Government. We will work together to ensure the public trust and establish a system of transparency, public participation, and collaboration. Openness will strengthen our democracy and promote efficiency and effectiveness in Government” (The White House, Office of the Press Secretary 2013).

The goal of this and other similar policies is to promote economic growth and to improve trust by opening government workings to scrutiny. As can be inferred from the executive order, the first commitment to building an efficient government is to open up access to its information and data that are used in decision-making. The technology community has also recognized the value of data in government transparency and citizen services innovation, and, as such, produced the Open Government Data Principles that state that government data should be (“The 8 Principles of Open Government Data” 2007):

1. Complete, i.e., data that are not subject to valid privacy, security, or privilege limitations must be made available.
2. Primary, i.e., data are published as collected at the source, not in aggregate or modified forms.
3. Timely, i.e., available as quickly as necessary to preserve the value of the data.
4. Accessible, i.e., available to the widest range of users for the widest range of purposes.
5. Machine Processable, i.e., reasonably structured to allow automated processing of it.
6. Non Discriminatory, i.e., available to anyone, with no requirement of registration.
7. Non Proprietary, i.e., available in a format over which no entity has exclusive control.
8. License-Free, i.e., not subject to any copyright, patent, trademark, or trade secret regulation.

The transition to a new view of data that acknowledges its entrepreneurial value and views security and privacy as crucial elements of federal government information occurred in 2016 when the US government formalized the commitment to data being managed as a strategic resource. The circular “Managing Federal Information as a Strategic Resource” stated that openness and interoperability increase operational efficiencies, reduce costs, and increase public access to valuable federal information and that making federal information discoverable, accessible, and usable “*can fuel entrepreneurship, innovation, and scientific discovery that improves the lives of Americans, and contributes significantly to national stability and prosperity*, and fosters public participation in Government” (Office of Management and

Budget (OMB) 2016, emphasis added). In this way, the turn from open and transparent to unlocking the value of data while ensuring its privacy and security began formally to take shape.

The shift, however, required some additional considerations of how data should be managed and handled, especially in the area of entrepreneurial activity, as this area includes a much larger group of actors viewing, using, and potentially exploiting the data. For example, the federal government should formalize the expectation of the entity *considering and protecting* an individual's privacy *throughout the data or information life cycle*. The choice of the term "lifecycle" is an important choice as privacy and data experts speak of lifecycle as a much broader conceptualization of data than storing data in a database.

The data lifecycle is a high-level, general framework that describes how data can flow through an organization or from the point of its collection or creation to dissemination and use. Like living organisms, data can be seen as being born (created) and then going through evolution (changes) up to its archiving or disposal. In general, the lifecycle includes such stages as collection, storage, sharing, analysis, preservation, and, if needed, deletion (Ball 2012). Viewing data through the lens of the lifecycle framework helps to maximize the benefit of data, minimize its cost, and improve its quality.

Lifecycle thinking helps to break down the processes of working with data into stages and understand what types of activities and resources are needed at each stage. As such, the database as a tool for storage and retrieval is a small portion of what must be considered. Moreover, managing throughout the lifecycle encompasses not only the data or information itself, but also the related resources, such as people, technology, finances, and so on (Kouper et al. 2021). The lifecycle approach is further amplified with explicit attention to privacy and security: "While security and privacy are independent and separate disciplines, they are closely related, and it is essential for agencies to take a *coordinated approach to identifying and managing security and privacy risks* and complying with applicable requirements" (Office of Management and Budget (OMB) 2016, emphasis added). While one can imagine coordinated to mean agencies working together to develop a unified approach, it also should be read as demanding agencies understand how to manage risks of aggregated data. For example, while an owner of one dataset may claim that individual names and addresses will not do harm and are fine to be collected and shared, once that information goes further through its lifecycle and becomes merged with other data, such as the data about police callouts, the result might cause massive harm. As such, "coordination" means entities working together to communicate risks and restrictions of an agency's own data and to understand how their datasets are part of a repository full of datasets.

Moreover, "The Federal Government shall provide members of the public with access to public information on Government websites. This responsibility includes taking *affirmative steps* to ensure and maximize the quality, objectivity, utility, and

integrity of Federal information prior to public dissemination, and *maintaining processes for addressing requests for correction of information disseminated publicly*" (Office of Management and Budget (OMB) 2016, emphasis added). The language in the quote envisions and places the burden upon the government entity as the owner of the website. Thus, the onus is on the government to monitor and ensure that the website meets the standards. In making decision in all these areas, "The design of information collections shall be consistent with the intended use of the information, and *the need for new information shall be balanced against the burden imposed* on the public, the cost of the collection, and any privacy risks" (Office of Management and Budget (OMB) 2016, emphasis added).

At the federal level, examination of the federal open government website can provide some insight into how these aspirations are accomplished and the challenges of managing data throughout the lifecycle. The data.gov website aggregates data from government publishers under the Open Government Data Act and the Federal Data Strategy ("About This Repository | Resources.Data.Gov" 2021). Its Data Governance Playbook acknowledges that effective data governance is foundational to leveraging data as a strategic asset and tasks every agency with identifying their data assets, developing a data management plan, and providing adequate data oversight and communications ("Federal Data Strategy: Data Governance Playbook" 2020). The task, however, is not easy to implement at the portal or repository level, where data comes from multiple sources and owners.

The data.gov website is a catalog that offers opportunities to download data from over 300,000 datasets in various formats. The catalog draws from federal, state, and local datasets, so it is a vast data resource that covers almost any aspect of government data gathering that can be imagined. There are seven broad categories, including agriculture, climate, and energy, and so on. In addition to hosting federal data, the repository also catalogs data from local governments, currently across ten cities, three counties, and eight states. Local governments, according to data.gov, should get their data on data.gov to make local data more discoverable and to show the international open data community how active the United States is in the domain of data (admin 2015).

The website brings together, organizes, and provides a search function for a broad array of datasets. For example, when examining the area of "older adult health," users of the website can view various datasets and can then click to access them. In the instance of dataset "Examination of Resident Abuse in Assisted Living Facilities in the United States, 2011 (ICPSR 34575)," the federal website links to a repository that provides the dataset; in this case, the data comes from the National Archive of Criminal Justice Data (NACJD), and the website restricts access to the dataset to users who have completed a Restricted Data Use Agreement and obtained approval from an Institutional Review Board (IRB), an entity that ensures ethical conduct of research.

This brief search and review of the data illustrates that the website serves an important function, as it is a single source of search for government-gathered and

government-funded data. As a catalog, though, data.gov is not responsible for the stages of data collection, description, analysis, and even, in most cases, preservation of data. It is a search and discovery interface that links to other websites. As such, one can infer that the rules of creating, accessing, and using the datasets are all controlled by the dataset hosts/sources. To assess the quality and utility of data, one must follow the links, visit the websites, and read about their procedures of data management and curation.

This design style of a centralized catalog seems to have been replicated in many other locations, across levels of government. Many US states have followed the federal example and implemented the open data initiatives, although they did so in their own way, leading to vastly different state websites. The policy landscape is also vastly different at the state level as less than half of the states, namely seventeen, have a solid public policy foundation (“State Open Data Laws and Policies” 2021).

For example, in 2014, the Illinois General Assembly unanimously passed a bill (Illinois, Public Act 098-0627) intended to increase transparency, accountability, and savings in government by establishing a new State Open Operating Standard. As a result, the State of Illinois built its Open Data Portal at [data.illinois.gov](http://data.illinois.gov), which, according to its *About* page, “lets you find data across state, find facts about your state, lets you create maps and graphs, and lets you freely download the data for your own analysis. Many of these datasets are updated daily, and some even more often.” The website provides instructions to those who are new to open data in a user guide, but the guide is only a basic description of how the site works and the technology behind it. Thus, the website uses CKAN, an open-source data management system that provides the functions of describing and storing resources (datasets), searching for them using keywords and categories (metadata), and registering and managing user accounts. As the guide points out, an account and login are usually not needed to search for and find data but is needed for other data lifecycle functions: datasets can be created, edited, etc. by users with the appropriate permissions.

As such, the website seems to create a three-tier system of interacting with data throughout its lifecycle. Anonymous use of the first tier allows one to view and sometimes download the data without registration. Creating an account enables the second tier with limited and likely specific use, such as uploading data. The third tier requires both registration and additional permissions, most likely, to edit other datasets and perform other functions. In addition, and similar to the federal website, the datasets are typically owned by the organization that created the dataset.

Each of these websites provide a rich set of examples that illustrate gaps in the basic design and decision-making that went into the creation of the “open” government data websites. These gaps point to the collective action dilemmas that can be conceptualized as lack of cooperation between individuals, communities, and government entities in the areas of open government data. The dilemmas concern the creation, dissemination, and use of the datasets.

For example, how did the dataset host determine the contents of the data and which portions of it should be open and widely available? Who or what government units were responsible for creating the data? Was there community engagement as part of this decision-making process? Moreover, if someone's data has been inappropriately included, is outdated, or otherwise is inaccurate or capable of misuse, how would individuals and/or others update or otherwise challenge the data? Finally, as open data concerns both the individual/community well-being and the functioning of civic society, what are the plans for long-term curation and preservation of these datasets, including their migration to newer technologies?

Both open data and open government are key for advancing government transparency and entrepreneurship, but the transparency does not seem to extend to decision-making in the creation and sharing of the datasets. Within the Illinois data portal, for example, dataset ownership is attributed to an organization, and it is mentioned that a portal may have more than one organization (State of Illinois Data Portal 2020). "Organization" here is merely a type of account with administrative rights that allows government departments to create and publish datasets and authorize others as contributors. It seems that individuals would need to find out who is behind such an account and interact with each organization that holds the original data (prior to it being added to the larger dataset), which is a short-sighted design decision as it demands individuals reach out to every individual organization to have data edited, updated, or removed.

The actual website provides no description of the mechanisms of data gathering, or the decision-making that goes into what data is "open" and how to challenge such a decision. There appears to be no mechanism for an individual to ask for the removal or correction of information. There is also no description of how the government can be made accountable for providing incorrect data. Lack of such information makes it unclear how to connect openness and participatory governance structures. On one hand, individuals "contribute" the data through their interactions with the government and its record-keeping; on the other hand, government compiles those records and, through digitization and aggregation, creates a new product. Data undergoes multiple transformations, and yet there is often limited citizen participation in the decision-making about the data itself or its aggregation.

The expectations of what is being done with data vary across the stages of data collection, sharing (publication) and access, and across the actors that are involved in managing data. If the community is not involved at the stages of data collection, but their data is now open for all to use, what are the implications for management and governance of this digital resource? And, finally, what are the obstacles to sustainable sharing and cooperation around government data? The remaining sections of the chapter take on these issues, beginning with a broad description of the knowledge commons governance models.

KNOWLEDGE COMMONS FRAMEWORK AND OPEN  
GOVERNMENT DATA

Aligning the general knowledge commons framework outlined in *Governing Knowledge Commons* (GKC; Frischmann, Madison, and Strandburg 2014) with a generic framework for open government data and data portals created as part of this movement generates insight into both frameworks. Doing so makes sense if one considers public data held by various levels of government as a public good generated by co-creation that yields public value (Foster and Iaione 2015; Jetzek, Avital, and Bjorn-Andersen 2019; McBride et al. 2019). As was mentioned earlier, data is an evolving entity that goes through various lifecycle stages, including creation, processing, analysis, dissemination, and even destruction. The data used in an OGD portal could be collected by and from individuals and separate organizations prior to being digitally aggregated. The digitization and aggregation create new knowledge resources, which can exacerbate the collective action dilemmas mentioned earlier. Following GKC terminology, we introduce three action areas that correspond to the three stages of the data lifecycle: (1) collecting the data, (2) publishing data as digitized aggregated material on a portal, and (3) accessing and using the data.

The first action arena, collecting data, encompasses collective action at the beginning of the data lifecycle. It involves multiple participants (actors), among them citizens and government officials, who have various levels of agency, responsibility, and power. Governments collect data as a means to keep records of government activities, such as notes from committee meetings and financial planning. Government also collects data as a function of being an organizing entity of citizens. In most of these situations, citizens are required to submit their data to a government for various reasons: registering to vote, securing a driver's license, purchasing or renting a property, etc. Government employees also play a role in the creation of data through research, observation, and statistics gathering (Boettcher and Dames 2018).

Within this action arena government entities have been collecting records and data for considerable periods of time. Many of these datasets existed in analog form and have been converted to digital form. This presents new challenges, especially in the area of citizen-generated data as data collected in analogue form might be presumed by those providing the information to have been less "open" and under specific rules of access. Data collection continues in parallel with the next action arena, data publishing, as governments continue requiring data from citizens and third parties create information from citizens' devices to make their own software more user-friendly. In case of the latter, data collection often occurs without the citizens' knowledge or consent.

The second action arena concerns the publishing, sharing, or dissemination stage of the data lifecycle, which could range from a simpler storage and discovery model



to more complicated models that involve processing, analytics, and visualization (Shah, Peristeras, and Magnisalis 2021). This arena requires an examination of actions around how and when to publish government data. First, it is important to recall that government-generated data comes, broadly, into the following categories: data generated by the governmental entity as a byproduct of its activities, data gathered by the government about its citizens and their activities and/or need to interact with the government entity, and data collected as part of the research and statistics efforts of the government. The latter category is the most documented area of actions, and it is mostly overseen by the federal government, so it is not examined in detail here.

For the first two categories, some local governments are trying to be active and transparent about how their publishing decisions are made, or at least they try to provide guidance for those who wish to publish data. For example, Nashville Open Data Portal contains links to resources that could be useful for data publishers and for the public that wishes to know how this local government makes data publishing decisions. The links, however, were defunct at the time of writing this chapter as they did not lead to any relevant materials, such as a data policy or open data documentation.<sup>1</sup>

Through open data policies that encourage citizen contributions and feedback, citizens could become involved in the decision-making process of what data should be published, how, and when. The existence of such policy creates a commitment to comply with the federal and state laws. Unfortunately, in many instances of creating an OGD portal, as briefly described earlier and explored further below, the design process seems to omit citizen cooperation and participation. Instead, legislators and government officials constitute the prime co-creators of such portals through incentivizing or enforcing portal creation efforts. Third parties, whether nonprofit or commercial, also become participants in OGD portal creation, as these organizations provide governments with the technological infrastructure to support these portals.

The third action arena involves data access and use. The protocols for accessing data differ according to whether the data is available in analog format or on a data portal. If the government data is only available from an analog source, those who request the data must often go in person to the appropriate office and fill out forms declaring who they are and why they need this information. There are thus greater obstacles that must be overcome to acquire these public records.

<sup>1</sup> Data policy for the Nashville open data portal has been archived by the Internet Archive and can be accessed at <https://web.archive.org/web/20210128125832/www.nashville.gov/Metro-Clerk/Legal-Resources/Executive-Orders/Mayor-Megan-Barry/mbo32.aspx>. Another version that is not linked from the portal can be found at [www.nashville.gov/departments/metro-clerk/legal-resources/executive-orders/mayor-megan-barry/mbo32](http://www.nashville.gov/departments/metro-clerk/legal-resources/executive-orders/mayor-megan-barry/mbo32). Since it is the same executive order by the mayor, most likely the link has become broken during a website update and was never repaired.

Digitally accessible data, such as the data on an OGD portal, is often much easier to acquire. Anyone can create an account, which they can use to access records that are available online. Such ease of use gives the impression that greater access to data means that greater numbers of citizens can participate in this action arena. This would be true, of course, if participation is measured only in terms of the numbers of people who access digitally accessible OGD and view or download the data. This narrow understanding of participation excludes creation and meaningful use of data. Disregarding meaningful use shifts the narrative of the OGD commons toward data access and dissemination (passive availability) rather than use (active co-production).

#### BLOOMINGTON OPEN DATA PORTAL

To further explore the issues discussed above, we use the city of Bloomington, Indiana, and its open data portal as a case study. As we consider open data portals to be an instantiation of digital commons, we assume that its design and governance would support cooperation and community participation and at least some forms of communal ownership, co-creation, and use. To test our assumptions, we use this specific case and apply the GKC framework and its concepts and guiding questions to understand the actions around the portal and their patterns and outcomes. Our approach involved a close examination of the portal (extensive search, browsing, reading, and note-taking of the portal materials until no new information could be inferred) and a data analysis of one dataset. For the latter, we downloaded the data and performed an exploratory analysis and visualization using R statistical packages.

We selected Bloomington for several reasons: it is an example of local government being committed to open data; it has a community of entrepreneurial, technology-driven actors; and its citizens are drawn from a large university environment, so they are somewhat predisposed to be proactive about data creation and use. Moreover, selecting a local OGD portal allows us to compare local initiatives with federal and state efforts. As was shown earlier, federal data websites often serve as aggregators of datasets, with data collection and ownership being distributed among many organizations. The city, on the other hand, gathers the most specific, local, and granular data, which makes it an important actor in government-driven data gathering. At the city level the rules of data collection, ownership, and access may be different. In this section we examine such local government actor and its data gathering: the Bloomington Open Data Portal B-Clear.

To accomplish this goal, we use the previously identified three action arenas as OGD moves throughout its lifecycle: (1) collecting data, (2) publishing data, and (3) accessing data. The GKC framework prompts us to ask questions about the background environment of this data commons, the attributes of this resource, governance strategies (including relevant action arenas) for managing the data commons, and patterns of interaction and outcomes that emerge from each action arena.

To build this case study, we first address the background environment and attributes of the resource. Second, we move to governance strategies, by examining a specific dataset, “rental properties” from B-Clear as an example of the types of granular attributes one might find in a data resource. Finally, we make observations about governance strategies and patterns of interaction and outcomes.

### *The City of Bloomington and the Open Data Portal*

The city of Bloomington, Indiana, is in south-central Indiana, roughly fifty miles southwest of Indianapolis. About 85,000+ people reside in Bloomington, including the students who attend Indiana University, making it the seventh largest city in Indiana by population numbers (Cubit 2021). The data collected and published by the city of Bloomington reflects these demographics and biophysical attributes. For instance, nuisance complaints and bus rides increase during the academic year, making relevant data a more prominent part of the portal homepage.

We consider the open government data portal in Bloomington to be a *type* of data commons. The background environment for this commons is comprised of three layers: (1) the city of Bloomington, (2) the city’s Information & Technology Services Department (ITS), which manages and oversees digital and smart initiatives, and (3) the city’s open data portal, B-Clear, which ITS created. Similar to the websites briefly described earlier, it is not apparent on the portal or in another easily accessible space what policies guide the creation of the portal, its structure, the definition of “open,” or the mechanisms that would allow citizens to challenge the “open” nature of data. Looking through the metadata of the datasets within the portal, one can determine the entity that uploaded various datasets, which is ITS administration’s individual employees. Thus, one can reasonably assume that the ITS implements policy, and ensures compliance with applicable law, but this is not explicitly stated anywhere.

Since the data portal is open, any member of the community can use it, including local politicians, city employees, citizens, and others who seek to ensure government is operating in a transparent manner. However, the use statistics or any other metrics of engagement are not part of the portal. One can assume, for example, that the city and its employees use some of the data to make decisions about city services and function of government. The data could also be used in an entrepreneurial manner by those who seek to do business with the city or desire to use city-driven data to build apps, provide services, or otherwise build business environments based on data-driven decision-making – for example, Lyft ride rentals and Lime scooters. The evidence of these uses did not appear in the searches that we performed. The only active use of the data that was supported by our examination was the analysis that local journalists have done on topics relevant to life in the city (Askins 2021).

*Open Data Portal and Its Data Resources*

B-Clear, which stands for “Bloomington Clear, or Be Clear about what we’re up to,” was launched in January 2016 at [data.bloomington.in.gov](http://data.bloomington.in.gov) as part of the mayor’s campaign to “open the doors of city hall, both figuratively and literally” (Banta 2016). The portal was envisioned to be a source that can inform and educate people about what is going on in city government. As the website explains: “B-Clear is a one-stop place to build an ever-growing assembly of useful data. We’re organizing it as open, accessible data so everyone can see and use it and manipulate it.”

It is important to note two distinct functions of the Bloomington Open Data Portal. First, the portal seeks to provide transparency and openness of governmental (and similarly situated) institutions, such as city finances, zoning commission, and others. And second, the portal opens access and use of data generated by government entities during the performance of their functions. This directly matches the federal commitments mentioned above, namely, to be a place of open data for the purposes of both governmental transparency and entrepreneurial use.

Yet this dual commitment creates difficulty in establishing boundaries and defining communities. Who is the producer of the data resources, the government or the individuals who supply their records? Similarly, the data can be used by many actors, including government employees, citizens, businesses, and so on. The absence of a well-defined user community and clear boundaries between the producers and the consumers poses a difficulty for defining the rules of access and sharing and evaluating contributions. On one hand, federal and state mandates require local governments to provide open data to citizens; on the other, citizens who use the data, especially for business and entrepreneurship, may become, in a way, “free riders” who pursue their selfish interests rather than cooperate for the public good. Citizens can also become “selfish” by withdrawing their information from the records pool and reducing the quality of aggregated data.

So far, this emerging conflict has been a potential rather than an actual conflict as it is not clear how much the data that is being published is actually used. Lack of understanding of the portal users and communities did not deter the city employees from publishing in the portal. As city departments pooled together their data resources, using the shared infrastructure of B-Clear, each city department was expected to contribute to the portal by making some data available. As of June 22, 2021, the B-Clear portal hosted 219 datasets, ranging from transportation to government finance.

The main page of the B-Clear website provides information about the datasets and the organizations contributing datasets (defined as CKAN Organizations or accounts discussed earlier). A “group” function allows for the grouping of datasets and information about the portal into categories. A separate google document linked from the website provides information about “what makes a good dataset” and how to upload data. Absence of user and community consideration is evident in the

absence of information for citizens whose data is captured in the various datasets. There is no explanation of the data shared, the authority determining publishing rules and restrictions and no means to request changes or challenge wrong or inaccurate data.

Moreover, an examination of the portal reveals that some datasets “available” on B-Clear are hosted or managed by third parties, adding a third layer of control and participation. For example, the CityProtect service links to an external website ([cityprotect.com.map](http://cityprotect.com.map)) that aggregates and creates visualizations of crime in Bloomington, IN. The visualization, with some level of granular data, is available without any barriers or agreement to terms of use, or other commitment to use of the visualization or the data contained, including locations of specific crimes. The information for visualization is updated regularly, using data gathered by the Bloomington police department.

### *A Specific Dataset: Rental Properties*

#### **Background**

Examining the history of one dataset that was made available through the B-Clear data portal allows us to understand the granular nature of a data resource in an OGD portal. The dataset is called “Rental Properties,” and it includes information about all properties that have been registered with the city of Bloomington as rental properties. The dataset was created in 2018 and is regularly updated. In April 2020, we downloaded the rentals dataset to conduct an exploratory analysis and examine what kind of information the dataset contains and what implications, both positive and negative, it may have for citizens.

The portal interface provides an activity stream for each dataset, which allows any person to see who uploads and updates the datasets. The rentals dataset was created by a user, whose account information does not provide many details. We know the name of the user, but there is no other information, such as their work position, job title, or contact information. At least five other people besides the author have been involved in updating the dataset. These people are primarily associated with the ITS department of the city of Bloomington, although some are also associated with Indiana University. A search on the Internet revealed that at some point these users were students from the university who worked as interns or participated in the “Summer of Code” program organized by the city.

#### **Dataset Attributes**

The dataset is available in a comma-separated (CSV) format. It contains twenty columns that provide information about properties, their owners, and agents who might be representing the owners, including the following: (1) property id, (2) date the property was registered, (3) date the rental registration permit expires, (4) address

of the property, (5) owner's name, (6) owner's address, including street address, city, state, and zip code, (7) name of the agent representing the owner, (8) agent's address, including street address and zip code, (9) the type of building where the rental property is, (10) type of the property itself, (11) number of bedrooms, and (12) maximum occupation allowance. Many columns (fourteen out twenty) have gaps in data, leading us to conclude that the rental property owners who filled out the forms described were probably not required to provide all the information. Anyone who visits the portal can download the full dataset and access all this information.

In what follows, we discuss the Rental Properties dataset in the context of the three action arenas: (1) collecting the data, (2) publishing the data, and (3) accessing the data.

### **Action Arena One: Collecting the Data**

In examining the action arenas, we seek to consider: "(1) participants (who may be either single individuals or corporate actors), (2) positions, (3) potential outcomes, (4) action-outcome linkages, (5) the control that participants exercise, (6) types of information generated, and (7) the costs and benefits assigned to actions and outcomes" (Ostrom 2005, 32–68). In the analysis here we focus on participants and discuss outcomes in a separate section below.

The participants in our first action arena, "collecting the data," are property owners who own rental property in Bloomington, the city of Bloomington, and the city's department of Housing and Neighborhood Development (HAND). Most of the information in the "Rental Properties" dataset is collected from property owners. These property owners, excluding Indiana University, are required by the city of Bloomington to register their property with HAND, which also inspects the property. The data is collected through the registration process, as every property owner has to fill out a registration form available on the city website. Thus, most of the information in the "Rental Properties" dataset is provided by property owners through the registration process.

To register, property owners use a fillable PDF form, with mandatory fields for the property address, owner's name and address, and some additional information about the property. Neither the form nor the website contains statements about whether the owner's information is intended to be publicly available. There is no description of mechanisms by which owners can request that this information not be shared. Thus, it seems that control over this information lies largely in the hands of the city.

Rental property procedures are governed by Title 16 of the Municipal Code, which states that the owner is required to register their rental unit (City of Bloomington IN 1998, Ordinance No. 21-32:16); such registration requires the owner to provide HAND with the name, address, phone number, email of the owner(s), and name of the in-state agent if the owner lives outside of Indiana. The title also contains a section "Disclosure" that prescribes the owner to keep this information

current. Title 16 does not contain any other statement regarding the sharing of the owner's name or address publicly or by request.

### **Action Arena Two: Publishing the Data**

The same groups of actors participate in the second action arena, "data publishing," but their actions concern the data that has already been collected and processed into a standardized digital format, e.g., a spreadsheet. At this stage of the data lifecycle, participants decide what data resources are published, when, how often it is updated, and who is responsible for data quality (i.e., accuracy, timeliness, and completeness). Guidance on these decisions is simply not found on the portal or any easily accessible city of Bloomington website.

It almost seems that both data producers and consumers are expected to be familiar with the various legal requirements and best practices that guide the decision-making, especially regarding critical and sensitive data. At the same time, as some of the data in this dataset may fit with the definition of "Protected Information," it should not be openly available, as, for example, is determined by the city of Nashville's TN Open Data Policy (City of Nashville TN 2016):

Data shall not be Open Data if it meets the definition of Protected Information. . . .  
 "Protected Information" means, but is not limited to (i) all confidential or restricted information, as defined in the Information Classification Policy, or any record or portion thereof, disclosure of which could be denied under the TPRA or other law; (ii) any record that contains a significant amount of confidential information . . . (iii) any record that reflects the internal deliberative or administrative process(es) of any Department . . . (iv) any record subject to privacy laws, or to copyright, patent, trademark or trade secret protection, or that are otherwise protected by law or contract.

While this is an incredibly broad and comprehensive policy, it is a perfect example of the wide-ranging concerns that can arise when publishing data in an open format. It also illustrates that open data policies are connected to other policies, such as the Information Classification Policy or the Tennessee Public Records Act (TPRA). Many publications contain data that can be personal or, in aggregate, can become highly sensitive. As such, one can argue that the widest of policies must exist when considering publication in an open environment. Such policies should not only define and regulate protected information, but also address the issues of potential harm from aggregation, loss of privacy, and terms that regulate user-generated content.

Moreover, it is important for policy-makers to understand that the citizens whose data is captured in the open data portal, sometimes long after the fact of providing the data to the city, may be surprised to discover that their data is shared in the portal environment. Much of the data "shared" was originally gathered in paper form. Aggregation of data in one place and ease of search and discovery are unintended uses of digitized record-keeping. As such, the decision should not be taken lightly.

Cost and effort of data publishing are two other considerations that need to be factored into OGD governance. Creating quality datasets is a time-consuming and often costly process, especially if the intended use is innovation, research, or other public benefit. The Rental Properties dataset is available in the spreadsheet format, which makes it easy to process with statistical packages. However, as will be shown later, it has missing data that affects its quality. Many of the other datasets on B-Clear are available as pdf documents. Such data is virtually useless and places the burden of conversion into the hands of those who seek to use the data. Without providing machine-readable data, the local government does not deliver on its second function – enabling analytical and entrepreneurial use.

### **Action Arena Three: Accessing the Data**

The participants in our third action arena, “accessing the data,” are prospective renters, HAND, and any citizens who might be interested in this information. From a review of the website, it appears that HAND maintains records of all rental properties, so one can reasonably assume they control the data source in this action arena. Yet it is not clear what entity converts the paper forms into a spreadsheet. Prospective renters can also visit the department and look at the property files of landlords before they sign a lease. Reviewing the records allows prospective renters to verify the information that landlords presented to them and make sure the latter comply with the necessary city codes and regulations, including safety and occupancy codes.

Prior to digitization, a renter would have to fill out a request for a public records form, such as the one in Figure 3.1, and provide their name and address.

Like all other public records, including vital records and criminal records, real estate and property records requests are reviewed by the city, which decides whether the request is valid, and the data is disclosable.

Data digitation changes the dynamics in this action arena. Rather than accessing individual records, interested citizens can download the whole dataset and review information on multiple property owners at once. Moreover, in addition to specific rental properties, one gets information about the property owner’s address and all properties they own. Such information potentially enables uses other than a prospective renter inquiring about a property of interest. We discuss some of these issues in the “Patterns and Outcomes” section.

The rental properties dataset was noted to have been updated as of June 27, 2021 and was available for download without registration or other restriction. Usually, services that offer online searches for public records, such as Doxpop, a service for accessing court records ([www.doxpop.com/prod/](http://www.doxpop.com/prod/)), require users to register before they can download data reports. Sharing digital records in an aggregated form as large datasets and making these datasets available for download without any registration, as B-Clear does, allows anyone to download the dataset and gain access not only to specific records of interest, such as a landlord of interest, but to all landlords and their personal identifiable information (PII).





**REQUEST FOR PUBLIC RECORDS**  
**CITY OF BLOOMINGTON, IN**

Please print:

Name of Person Requesting Records: \_\_\_\_\_

Address: \_\_\_\_\_

Contact Information:      e-mail \_\_\_\_\_  
   phone \_\_\_\_\_  
   fax         \_\_\_\_\_

Records Requested. Please, be specific. You may use the back of form if necessary.

[Large empty rectangular area for recording requested records]

This request is,       for permission to inspect records.  
                                  to request a copy of records. *Please note there may be a fee for copies.*  
                                  check here if you want to be told about that fee before copies are made

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Date

FIGURE 3.1. Public record request form

The downloader does not have to provide any information about themselves, and HAND makes no determination regarding their ability to disclose the information. This situation suggests that while HAND maintains the records regarding rental properties, it (1) no longer directly interacts with the prospective renter and (2) no longer provides governance regarding the ability to disclose the information. In an OGD portal, control of this information seems to have transferred to the user of the information. There is no governance structure around this or other B-Clear datasets once it gets digitized and aggregated, and the user is free to use the resource however they wish.

**Patterns and Outcomes**

Like other data resources on the portal, the Rental Properties dataset provides a glimpse into city life. In this case, we can learn, for example, how many properties

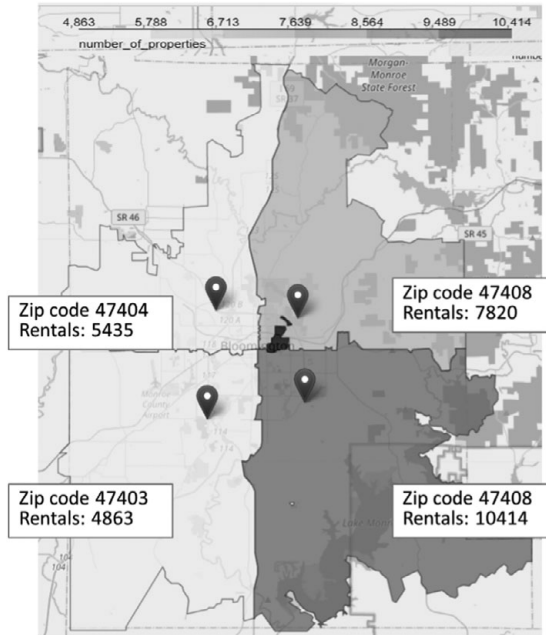


FIGURE 3.2. Number of rental properties by zip code in Bloomington, Indiana

are registered as rentals in Bloomington (28,739) and where the owners come from (21,888 from Indiana, 3,938 from Illinois, and 756 from California).

The dataset may provide useful insights for residents, visitors, or business owners about the geographical distribution or density of rental properties. By geocoding the property addresses, one can create a map that shows which zip code areas have a higher concentration of rental properties. As Figure 3.2 shows, the most concentrated area of rentals is at the bottom right corner of Bloomington with the zip code 47408. Not surprisingly, Indiana University is located near that area, and the students who attend the university are one of the largest groups of people that rent houses and apartments in Bloomington.

Aggregate mapping, such as discussed earlier, represents a positive implication of having rental property data available. More disturbing implications may come from the fact that the dataset contains personal identifiable information about the property owners, including their names and addresses.

While many landlords in the dataset are companies, many are also individual owners. By analyzing the dataset one can identify the biggest individual owners in Bloomington as well as learn about small property owners, where they live, and where their properties are located. We discovered that while several individuals own hundreds or thousands of properties in Bloomington, the majority of owners are small-scale landlords who own between one and five single-unit houses, condos, or apartments.

The detailed PII information provided in this dataset creates multiple risks for individual owners. The name and address data can be combined with other data and analyzed in a way that increases the risk of exposing other PII, such as gender, race, workplace, family members, and other descriptors). Learning about their data being exposed in this way, rental property owners may not trust the city government and be less willing to provide data to it. One pattern that we see is that some owners try to protect themselves and their information by providing a PO Box address rather than their street address. This may be because they know they can no longer exert control over their information once it is in the city's databases, and they may not agree with the city's governance of such highly sensitive personal information. Rental property owners may be viewed as wealthy and become targets for crimes, especially when their home addresses are publicly available.

### *Governance of the Bloomington Open Data Portal*

Our analysis of the datasets and how they are presented on B-Clear reveals that the city of Bloomington at some point identified open data as one of their priorities, requiring the city's departments to share their data to the portal. Historically, some data has policies that govern its sharing, such as city budgets and spending but other data either has no such policies or has policies that were designed for the analog record-keeping based on the "walk-in" scenario for requesting the data. A citizen who is interested in the data walks into the city hall and submits the request. A clerk provides a specific record or records that have been requested and no other data is made available.

Governance of open data on B-Clear exists in bits and pieces rather than as a systematic set of rules and policies that all participants have developed and approved. For example, for the "City of Bloomington Annual Compensation" dataset, there is the following statement: "I agree, pursuant to Indiana Code 5-14-3-3(f), that any information, including the names and addresses of government employees, that I obtain by viewing, printing and/or downloading will not be used for commercial or political purposes." For some datasets, such as "Cities and Towns Report," the metadata provides information about the reasons this dataset is made publicly available (although the latest data is from 2015): "Per IC 5-11-1-4 every municipality and local government is required to provide electronically and, in a manner prescribed by the state examiner, financial reports for the fiscal year not later than sixty days after the close of the fiscal year."

Yet the majority of datasets contain no restrictions whatsoever. For example, the rental property data, the guns reported stolen, the domestic battery data, the hate crimes data, and the SarsCoV2 Data 2020, to name a few, are all available without restriction or limitation. The "About" page on the B-Clear portal contains several sentences that declare B-Clear a "one-stop place" for an ever-growing assembly of useful data. The sentences capture the city of Bloomington's priorities of openness,

but one can argue the governance of the portal and the Bloomington Open Data website are incredibly lacking, especially in terms of transparency of process, accountability, and citizen engagement in the process. Moreover, the use of the phrase “our data” on that page is troubling as it seems to be inclusive of individual record information owners, but both the term and its underlying concept (referent) must be challenged and redefined in the context of citizen participation and contributions in data gathering, curation, and publishing.

### *Data Dump without a Lesson Learned*

In 2018, Bloomington began tracking opioid overdose deaths online through the portal and created a visualization map that pinpointed, very precisely, the addresses of overdose victims within city limits. After a loud outcry by many in the community, the city of Bloomington altered the dataset to remove “specific addresses” and justified the existence of the dataset by stating:

Sharing this public data about where these deaths are occurring can help give those working toward a solution more tools to help those who are suffering . . . At the same time, we understand that this is a sensitive issue. As we have done since launching the site, we will continue to consult with those on the front lines of this crisis to follow best practices. (Axelrod 2018)

The actions and the following statement are another example of the lack of governance-oriented thinking that appears in the B-Clear Portal. If the data needs to be made available to those who are working toward solutions, it does not have to be open to everyone without restriction. For example, the city could restrict the open publication either to those who registered or to those whose accounts have specific permissions. The city could even require entities to specifically request access with a permissible use explanation. None of this was done. Similar to the rental properties data, the city seems to have decided to act in the spirit of complete openness without fully considering the consequences of its actions for all participants.

This data upload had a public reaction, but many other datasets that have not yet been combed through, integrated with other data, and visualized, remain to be open for any use or misuse. As the inference is that citizens are given no input at the earlier stages of the data lifecycle, they may find they will be unprotected in the future from aggregated datasets.

## CONCLUSION

Like many communities across the United States, Bloomington seeks to adhere to the federal commitment of providing “the public with knowledge of the Government, and to ensure the accountability of Government” and to maintain the public trust by establishing a system of transparency, public participation, and

collaboration. While a lot of knowledge in the form of open data has been provided, one can argue, the later parts of this commitment, namely, public participation and collaboration, have been lost in many communities.

As the chapter has briefly highlighted, some communities have embraced collaboration and trust as essential elements of “open” data, government or otherwise. But it appears more needs to be done. Viewing government data about the citizens as a digital data commons is one important step in the right direction that would allow open data portals to codify their policies and collect, publish, and access data with a purpose. It would allow local governments to prioritize which data is most important for which community and how the public and businesses can best use government data as a shared resource. If trust, participation, and accountability are the hallmarks of the government commitment to open data, then they must be embedded as implementing principles of the data portal creation, maintenance, and governance.

Of course, this is an important first step, but more knowledge must be brought into the data that exists on the data portal as well. Digitized data, both when converted and when created in digital form, is within an infrastructure (portal) that enables sharing, aggregation, and use of additional services. Every decision about all of the data must be made with this in mind. As local governments get involved in data aggregation and sharing, they cannot ignore the fact that they, the citizens, and all other participants become part of a global digital platform that is more and more pervasive in everyday life.

Viewing OGD as digital data commons will also enable local governments to collaborate with other stakeholders and draw on the expertise of legal and other scholars who have been studying governance of knowledge and other resources. Multiple governance models have been discussed by scholars that prioritize participation, innovation, or control (Janowski, Estevez, and Baguma 2018). Exploring these models in the context of OGD will enrich both the policy debates and the collective action around digital platforms and data commons. It will help to create mechanisms of safe and useful releases of open government data without encouraging selfish or harmful behavior (Yakowitz 2011). Finally, none of this works without full citizen participation during the entire data lifecycle, including decision-making around user or citizen-generated content and the ability to access and use data. A digital data commons framework will help the governance become clearer as it will engage the citizens and government to work together in a way that benefits all.

#### ACKNOWLEDGMENT

The authors would like to thank the Ostrom Workshop and the Weimer Faculty Fellows program for funding the research within this study. The authors also wish to thank Hilda Hadan and Jaime Carini, doctoral students at Indiana University Bloomington, who were involved with early data gathering and research in this study.

## REFERENCES

- “About This Repository | Resources.Data.Gov.” 2021. <https://resources.data.gov/about/about-this-repository/>.
- admin. 2015. “Get Your Local Government on Data.Gov.” [www.data.gov/opendata/get-local-government-data-gov/](http://www.data.gov/opendata/get-local-government-data-gov/).
- Askins, Dave. 2021. “Bloomington’s Dozen Water Main Breaks in a Week Make for Record-Setting Start to 2021.” *The B Square*. <https://bsquarebulletin.com/2021/02/26/a-dozen-water-main-break-in-a-week-adds-to-bloomingtons-record-setting-start-to-2021/>.
- Attard, Judie, Fabrizio Orlandi, Simon Scerri, and Sören Auer. 2015. “A Systematic Review of Open Government Data Initiatives.” *Government Information Quarterly* 32 (4): 399–418. <https://doi.org/10.1016/j.giq.2015.07.006>.
- Axelrod, Jason. 2018. “Indiana City’s Controversial Online Overdose Map Modified Following Numerous Complaints.” *American City and County*. [www.americancityandcounty.com/2018/07/16/indiana-citys-controversial-online-overdose-map-modified-following-numerous-complaints/](http://www.americancityandcounty.com/2018/07/16/indiana-citys-controversial-online-overdose-map-modified-following-numerous-complaints/).
- Ball, Alex. 2012. “Review of Data Management Lifecycle Models.” <http://opus.bath.ac.uk/28587/1/redmirep120110ab10.pdf>.
- Banta, Megan. 2016. “Mayor Wants to ‘B Clear’ with New City Website Function.” *The Herald Times*, [www.heraldtimesonline.com/story/news/local/2016/01/06/mayor-wants-to-b-clear-with-new-city-website-function/47897609/](http://www.heraldtimesonline.com/story/news/local/2016/01/06/mayor-wants-to-b-clear-with-new-city-website-function/47897609/).
- Boettcher, Jennifer C., and K. Matthew Dames. 2018. “Government Data as Intellectual Property: Is Public Domain the Same as Open Access?” *Online Searcher* 42(4): 43–48.
- City of Bloomington IN. 1998. *Title 16. Residential Rental Unit and Lodging Establishment Inspection Program. Municipal Code*. Vol. Ordinance No. 21-32. [https://library.municode.com/in/bloomington/codes/code\\_of\\_ordinances](https://library.municode.com/in/bloomington/codes/code_of_ordinances).
- City of Nashville TN. 2016. *Executive Order Number 032 Open Data Policy*. [www.nashville.gov/departments/metro-clerk/legal-resources/executive-orders/mayor-megan-barry/mbo32](http://www.nashville.gov/departments/metro-clerk/legal-resources/executive-orders/mayor-megan-barry/mbo32).
- Cubit. 2021. “Indiana Cities by Population.” [www.indiana-demographics.com/cities\\_by\\_population](http://www.indiana-demographics.com/cities_by_population).
- Digital Public Goods Alliance. n.d. “Digital Public Goods Standard.” <https://digitalpublicgoods.net/standard/>.
- “Federal Data Strategy: Data Governance Playbook.” 2020. <https://resources.data.gov/assets/documents/fds-data-governance-playbook.pdf>.
- Foster, Sheila, and Christian Iaione. 2015. “The City as a Commons.” *Yale Law and Policy Review* 281. <https://doi.org/10.2139/ssrn.2653084>.
- Frischmann, Brett M., Michael J. Madison, and Katherine J. Strandburg, eds. 2014. *Governing Knowledge Commons*. Oxford: Oxford University Press.
- Greenstein, Shane. 2013. “Digital Public Goods.” *IEEE Micro* 33 (5): 62–63. <https://doi.org/10.1109/MM.2013.96>.
- Janowski, Tomasz, Elsa Estevez, and Rehema Baguma. 2018. “Platform Governance for Sustainable Development: Reshaping Citizen-Administration Relationships in the Digital Age.” *Government Information Quarterly* 35 (4): S1–16.
- Jetzek, Thorhildur, Michel Avital, and Niels Bjorn-Andersen. 2019. “The Sustainable Value of Open Government Data.” *Journal of the Association for Information Systems* 20 (6). <https://doi.org/10.17705/ijais.00549>.
- Kouper, Inna, Karen Tucker, Kevin Tharp, Mary Ellen van Booven, and Ashley Clark. 2021. “Active Curation of Large Longitudinal Surveys: A Case Study.” *Journal of ESience Librarianship* 10 (3). <https://doi.org/10.7191/jeslib.2021.1210>.

- McBride, Keegan, Gerli Aavik, Maarja Toots, Tarmo Kalvet, and Robert Krimmer. 2019. "How Does Open Government Data Driven Co-Creation Occur? Six Factors and a 'Perfect Storm'." *Government Information Quarterly* 36 (1): 88–97. <https://doi.org/10.1016/j.giq.2018.11.006>.
- Office of Management and Budget (OMB). 2016. "Circular No. A-130. Managing Federal Information as a Strategic Resource." <https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/OMB/circulars/a130/a130revised.pdf>.
- Ostrom, Elinor. 2005. *Understanding Institutional Diversity*. Princeton, NJ: Princeton University Press.
- Schudson, Michael. 2018. *The Rise of the Right to Know*. Cambridge, MA: Harvard University Press.
- Shah, Syed Ifthikhar Hussain, Vassilios Peristeras, and Ioannis Magnisalis. 2021. "DaLiF: A Data Lifecycle Framework for Data-Driven Governments." *Journal of Big Data* 8 (1): 89. <https://doi.org/10.1186/s40537-021-00481-3>.
- Siddarth, Divya, and E. Glen Weyl. 2021. "Technology Governance: The Case for the Digital Commons." World Economic Forum. [www.weforum.org/agenda/2021/06/the-case-for-the-digital-commons/](http://www.weforum.org/agenda/2021/06/the-case-for-the-digital-commons/).
- State of Illinois Data Portal. 2020. "User Guide – Data.Illinois.Gov." <https://data.illinois.gov/pages/guide>.
- "State Open Data Laws and Policies." 2021. *National Conference of State Legislatures*. [www.ncsl.org/research/telecommunications-and-information-technology/state-open-data-laws-and-policies.aspx](http://www.ncsl.org/research/telecommunications-and-information-technology/state-open-data-laws-and-policies.aspx).
- "The 8 Principles of Open Government Data." 2007. <https://opengovdata.org/>.
- The White House, Office of the Press Secretary. 2013. "Executive Order – Making Open and Machine Readable the New Default for Government Information." <https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government>.
- Yakowitz, Jane. 2011. "Tragedy of the Data Commons." *Harvard Journal of Law & Technology* 25 (1): 1–67.