

One of the goals of opening government data is to improve governance by expanding information and access in ways that draw new actors, interests, and influence into government decision making. We use the concept of an information polity as a heuristic device to understand the ways that the new mixture of stakeholder interests are shaping, and are shaped by, new information flows and technologies (Bellamy and Taylor 1992, 1998).

The value in thinking about an open data initiative in terms of information polity components is that it can help agencies identify and distinguish important data sources, flows, and stakeholders. This is an important starting point for understanding how the creation or modification of data sources, flows, or governance relationships affects the interests of various stakeholders and ultimately impacts value creation.

An Information Polity is the collection of stakeholders, data sources, data resources information flows, and governance relationships involved in the provision and use of government-held and nongovernmental data sources.

There is growing interest at all levels of government to increase access to and use of government data in support of good governance. As a result, public agencies are under pressure to create new capabilities to achieve this goal. A common assumption when opening government data is that simply supplying more data freely and in more formats will lead to more use. That use will lead to value creation and, in turn, will motivate government to make the necessary changes to continue opening more data. But, we know from experience, that supplying more and more data does not necessarily produce the results we anticipated.

Figure 1. View of an information polity.

Figure 1. View of an information polity.

Data sources and flows. There are many different kinds and sources of information that are important for improving government, but not all kinds or sources of data are 'held' by government. When thinking about the value of open data initiatives, it is expected that government data will be combined with other data sources. While government has gotten better at sharing data between government agencies and levels, we still do not have a good idea of how to share data with 'the public,' 'app developers,' or others outside the traditional boundaries of government information sharing. Nor do public leaders have a good understanding of the implications of combining government data sources and secondary data sources. This is an important analytical distinction because it allows them to simultaneously think about the governance of 'government-held' data (such as datasets on a site like Data.gov) and 'non-governmental sources' of data (such as comments, geo-coded data, and social network data on third party social networking sites like Facebook or Twitter).

We present open data initiatives as potentially having a primary data source and a secondary data source; both collected as a result of the interactions in the information environment. The data flows of Primary Data Sources are shown by the light blue arrows and the data flows of Secondary Data Sources are shown by yellow arrows.

In our description, government is the Primary Data Source, Provider, and Data Resources. The Primary Data Source sends the data to the Primary Data Resource governed by the Primary Provider, which is then made available to the Primary Audience/User. The simple picture considers the possibility of the Primary Audience/User obtaining some of the same kinds of data directly from the Information Environment themselves. An example would be a citizen observing directly the sanitary conditions of a restaurant. However, there are aspects of the Information Environment that cannot be observed directly by the Audience/User, such as taking the temperature of a freezer in a restaurant that stores food.

Primary Components of an Information Polity.

- Information Environment is the multiple contexts from which data is extracted, encoded, and otherwise made visible
- Primary Data Sources, in our view, are the public employees that interact with the information environment and encode the original data required as part of a government program, process, or reporting requirement. This data is then entered into a government information system.
- Primary Data Resources are the access/ interface tools combined with primary data sources that are created that provide users with the data. The resource includes the data files, the software, networks, platforms, and organizational arrangements needed (such as creating a website or an application).
- Primary Data Providers are the government agencies with the authority and responsibility for creating and maintaining the primary data sources and resources. In this role, government providers acquire and structure data files, create requisite policy, governance, and management arrangements necessary to establish and maintain the primary data resource.

- Primary Audiences/Users are the persons or groups who are part of the government program, process, or reporting requirement (inside or outside of government) that are the intended users. The government data source and resource were created for them to advance some government objectives.

In a Web 2.0 world, Secondary Data Sources, Providers, and Resources represent the explosion of new stakeholders (e.g. app developers, social networking sites, or citizens). Secondary Data Sources can also interact with the Information Environment to create data and initiate additional secondary data flows. The Secondary Data Resources extract Secondary Data Sources, which are managed and governed by the Secondary Providers. Data from Secondary Data Resources can go directly to Secondary Audiences/Users, or be combined with Primary Data Sources and flow to the Primary or Secondary Audience/Users. Of course multiple Secondary Data Sources and Resources are possible, but are omitted from the figure for simplicity and readability.

Identifying the data sources and flows highlights the potential challenges in sharing and integrating data sources. Data does not exist in the wild; it is deliberately created by socio-technical processes. These processes may be as straightforward such as a digital temperature sensor sending a reading to a weather database or a nurse taking a blood pressure reading on a patient and entering it in a medical record. The processes may also be much more complex, such as a psychiatric emergency room physician deciding whether to identify a patient as a risk to himself or others and recording that in a medical record or a citizen providing analysis and comments on a piece of pending legislation. The usability of data, or its fitness for use, depends in large part on the nature of the encoding processes and data management practices.

Likewise, even within the same government department, different units can have widely varying data definitions, standards, and encoding methods for similar kinds of data. As long as they act independently, these differences are not necessarily an issue; however, when integrating or comparing data, the differences may have serious consequences that require governance decisions. These kinds of issues can become more problematic when data from non-governmental sources is involved. The more data providers are involved in working with a data resource, the more complex the roles and relationships.

Actors and roles. Government practitioners working to open government data make up only a part of the larger group of stakeholders who have an interest in and ability to influence how data is acquired, accessed, and used. Stakeholders can influence events in an information polity. Their presence represents an understanding that opening data directly involves the internal management of agency and information technology systems as well as external stakeholders that are producers or consumers of data. Other stakeholders represent other sources of influence on action, connected to the social and political environment, the nature of the information of interest, and the institutional context of law and policy. Open data initiatives disrupt government's traditional role as 'holder' or 'owner' of the data. In thinking about open data governance, we need to re-think government's role in relation to the entire set of new stakeholders. One possibility is to characterize government, as well as all other stakeholders as stewards (Dawes, 2010) of primary and secondary data sources. The idea of stewardship focuses on the joint responsibility of all public officials and government organizations, as well as other stakeholders to assure the accuracy, validity, security, management, and preservation of data holdings. It demands that primary and secondary data be acquired, used, and managed as a resource that has organizational, jurisdictional, or societal value across purposes and over time (Dawes, 1996).

Secondary Components of an Information Polity.

- Secondary Audiences/ Users are the persons or groups that want access to or use the primary data in ways other than 'originally' intended.
- Secondary Data Sources are data that comes from sources other than the government provider. The data may be social media comments, sensor data, or other types of information collected from the information environment or from users directly.
- Secondary Data Resources are the access/interface tools that are created that provide users with the data. These resources draw from both secondary data sources as well as primary data sources. The resources include the data files, the software, networks, platforms, facilities, and organizational arrangements needed (such as creating a mobile app).
- Primary Data Providers are the government agencies with the authority and responsibility for creating and maintaining the primary data sources and resources. In this role, government providers acquire and structure data files, create requisite policy, governance, and management arrangements necessary to establish and maintain the primary data resource.
- Secondary Data Providers are the persons or groups that acquire the data from the government or secondary data sources and redistribute it in a modified way that provides benefits or additional impacts beyond those resulting from access to the original data resource.

Governance relationships Stakeholders in an information polity are oriented toward steering (consensually or antagonistically) (Corry, 2010) the data sources and resources involved in an open data initiative. Each stakeholder has interests in the nature of and success of an open data initiative, such as enhanced program effectiveness for a particular program or increased political influence or national security. The governance of open data initiatives involves creating policies, business processes, social processes, technologies, standards, meaning and interpretation, and adding value.

As a Primary Provider, a government agency can have governance relationships with the Primary Audience/Users, with the Data Source, Secondary Data Provider, and with the entities in the Information Environment. We show the governance relationship between the Primary Provider and the Primary Audience as a reciprocal one. The Primary Provider can influence the Primary Audience directly through data provision, and through other incentives, sanctions, and persuasive methods. Similarly, the Primary Audience/User can influence the Primary Provider through political processes and direct participation in decision making and data use.

Primary & Secondary Audiences/Users are not a monolithic entity, but an aggregate of persons and groups with various interests in access to and use of the data. These interests may not be fully aligned, leading to conflict between stakeholders and the introduction of competing goals for Primary Providers. The influence mechanisms used by Primary and Secondary Audiences/Users can take many