

Valuing public data

Last Updated November 26, 2018

Public data are collected because they are essential to government function. Whether or not they are freely available to the general public varies between governments. Here we explore what public data are and the implications of charging for data in terms of cost recovery and impacts on data usage.

Just as the supply of basic physical infrastructure – power, transport, telecommunications – is essential to the traditional economy, so the supply of basic information ‘infrastructure’ – geography, weather, transport, etc. – is essential to the ‘information’ economy.

– [Rufus Pollock](#)

What are public data?

Public data are collected by federal, state, and local governments. Common public data include geospatial (parcels and roads), demographic (census), meteorological, and hydrologic data. Public data of particular interest to the Internet of Water are those *non-personal* data relevant to creating a [water budget](#) (quality, quantity, and use). The importance of public data, and the infrastructure to deliver those data, continue to grow with the rapid development of an ‘information’ economy. And yet, limited funding has made it difficult for public agencies to develop and maintain a robust data infrastructure to keep pace with technology. Technology has profoundly changed how data are collected and disseminated (Table 1). To maximize data usage and impact, there must be investment in the underlying data infrastructure.

Table 1: Paradigm shift from a traditional to an information-based economy (adapted from Uhler *et al.* 2015).

Traditional Economy (Hard-Copy)	Information Economy (Digital)
Data are fixed & static	Data are transformative & interactive
Data are physical, hard copies	Data are digital, software
Digital tools cannot manipulate content	Digital tools available online
Data are local and hard to access	Data are global and easy to access
Limited types of data	Unlimited types of data
Making copies is cumbersome & introduces errors	Easy to copy & distribute data
Significant marginal cost to distribute	Near zero marginal distribution costs
Centralized production of data	Distributed & integrated data production is possible

Public data are collected because they are essential for government function. For instance, the National Weather Service collects and disseminates weather data to protect lives and property as well as to benefit the economy. While the essential nature of the data requires public sector involvement, there is wide variation in government policies regarding how those data may be re-used by the general public and private industry.

To charge or not to charge for data

Secondary demand for public data exists because these data can be used to understand government, infrastructure, society, economy, nature, and so on. In the United States, public data are freely available by [executive order](#). Arguments favoring this approach include:

- 💧 Greater potential societal value is possible with greater usage.
- 💧 Public data are a public good because taxes were used to pay for data collection. The public should not have to pay a second time to access data.
- 💧 Democratic values and transparency of the government would be undermined by restricting access.

However, not all governments or public agencies share the same perspective. Many countries in Europe attempt to recover costs by charging for data access. In the United States, public agencies have tried, and continue to consider, charging for data access. Real-world experiments demonstrate the impact of charging for public data in terms of data usage and cost recovery.

Free data increases usage

Free data are used more. There are repeated instances where organizations have ceased charging for data and have seen an increase in data usage. The U.S. Geological Survey (USGS) saw a 43% increase in new users after they stopped charging for access to Landsat imagery in 2008. The number of downloads doubled from the Australia Bureau of Statistics when data were made freely available in 2005.

Conversely, when a charge is associated with data, usage decreases. Several studies have compared the U.S. (free data access) with Europe (charge for data access) and found that public data in the U.S. has a benefit of \$39 for every \$1 spent, compared with a benefit of \$7 in Europe (Table 2). The challenge is that increased usage and value creation does not necessarily translate to revenue support for these public agencies. With tighter budgets, many agencies are attempting to find ways to recover costs.

Table 2. Estimated cost and benefit of public data in Europe and the United States.

Category	Europe (\$B)	United States (\$B)	US : Europe
Cost (Investment)	11	22	2.0
Benefit (Value)	78	855	11.0
Benefit : Cost	7.1	38.9	

Cost recovery is difficult

Restricting data access creates substantial administrative overhead, as well as hidden costs from lost opportunities, barriers to innovation, and sub-optimal data quality from limited use and accountability. One of the primary challenges is that public agencies incur costs to collect and disseminate data while secondary users receive most of the benefits. For instance, the Australian Bureau of Statistics lost \$3.5M annually by making data freely available while producing between \$6M and \$25M annually in value for secondary users.

In general, cost recovery efforts for data have been met with limited success and unintended consequences because:

- ◆ Secondary demand is not large enough to support cost recovery. For instance, one fisheries data program recovered less than 1% of costs by charging for access.
- ◆ Charging other government users merely shifts expenses between agencies rather than recovering government costs.
- ◆ [Data's unique attributes](#) often result in market failures.
- ◆ High prices (full cost recovery) lead to predatory and anticompetitive practices, such as the creation of government-owned corporations that exclude others from public data. Finland and Sweden provide examples of government having a competitive edge by using high-quality data to create better products and only making low-quality data publicly available. To avoid such pitfalls, government should separate the provision of public data from value added (commercial) products.
- ◆ There is a growing belief that data should be free and many refuse to pay for data, instead searching for free data from other sources or simply not using data for decision-making.

Balancing costs and accessibility

While there is a perception that data should be free, there is no such thing as free data. Water is free, but it costs money to collect, treat, and distribute water to users. Similarly, public data may be free, but it costs money to collect, manage, and disseminate data to users. The U.S. has an executive order to make data freely available, but public data may not reach their potential due to inadequate funding. Inadequate funding may result in low data usage because the data are not very discoverable, accessible, or usable. The [Internet of Water](#) aims to improve data infrastructure so that public water data moves from a market with low activity (sleeping) to an active market (dynamic) (Figure 1).

Instances where governments attempt to recover costs by charging for data access may result in lower usage because of the cost to secondary users (sleeping). Another possibility is that the costs are so prohibitive that public sector monopolies form, making limited use of the data relative to having many data users. A third option is that private sector markets emerge that collect their own data and forgo public data. None of these outcomes are ideal for water resources, particularly when addressing regional challenges that have formed over decades and cross multiple jurisdictions.

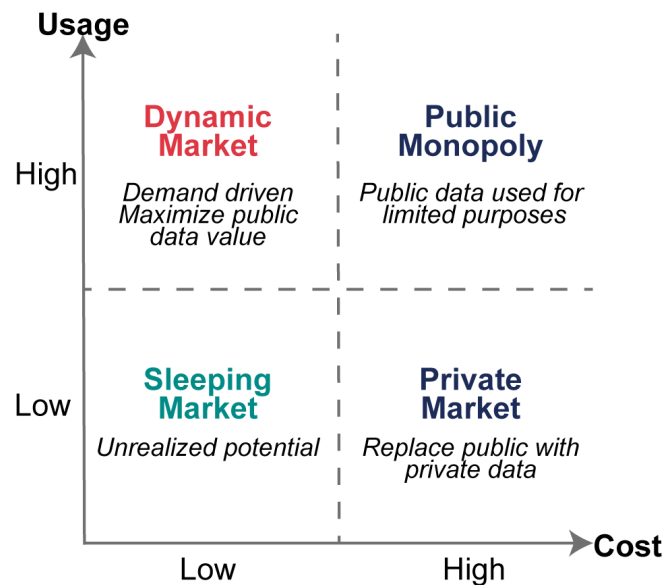


Figure 1: Markets that are likely to form based on cost and usage (impacted by discoverability, accessibility, and interoperability). Public data in the U.S. are typically free and will be on the spectrum between a sleeping and dynamic market.

For more information:

GEO. 2015. [The Value of Open Data Sharing](#).

Houghton. 2011. [Costs and Benefits of Data Provision](#). Report to the Australian National Data Service.

Mayo & Steinberg. 2007. [The Power of Information](#). Cabinet Office Minister of Policy Review UK.

Miller *et al.* 2013. [Users, Uses, and Value of Landsat Satellite Imagery – Results from the 2012 Survey of Users](#).

Pira International Ltd. 2000. [Commercial Exploitation of Europe’s Public Sector Information](#).

Pollock. 2008. [The Economics of Public Sector Information](#). University of Cambridge.

Uhlir *et al.* 2015. [The Value of Open Data and Sharing](#). Report by GEO and CODATA.

Uhlir. 2009. [The Socioeconomic Effects of Public Sector Information on Digital Networks: Towards a Better Understanding of Different Access and Reuse Policies: Workshop Summary](#). National Academies Press.

Uhlir & Schröder. 2007. [Open Data for Global Science](#). *Data Science Journal*

Vickery. 2010. [Review of Recent Studies on PSI Re-Use and Related Market Developments](#).

Weiss. 2010. [Borders in Cyberspace: Conflicting Public Sector Information Policies and Their Economic Impacts](#). The National Academies Press.