



# Fostering Innovation, Creating Jobs, Driving Better Decisions: *The Value of Government Data*

*Prepared by*

**U.S. DEPARTMENT OF COMMERCE  
ECONOMICS AND STATISTICS ADMINISTRATION**

**JULY 2014**



Fostering Innovation, Creating Jobs, Driving  
Better Decisions:  
*The Value of Government Data*



Economics and Statistics Administration  
Office of the Chief Economist



July 2014



# Fostering Innovation, Creating Jobs, Driving Better Decisions: *The Value of Government Data*

## Foreword

In March, The Department of Commerce unveiled the *America is Open for Business: Strategic Plan, Fiscal Years 2014-2018*. One of the plan's five priority areas is to redefine how we manage, optimize, and enable public access to our treasure trove of data. Specifically, the plan pledges to improve government, business, and community decisions and knowledge by transforming Department of Commerce data capabilities and supporting a data-enabled economy.

The Department's efforts are part of a broader Open Data Initiative, which has embraced innovation as a means of further unlocking the value of Government data. The initiative focuses on making new data available to the public while rigorously protecting privacy; putting data into forms that are more accessible and usable; and making entrepreneurs and innovators aware of Government data and how to access it.

Here at the Department of Commerce, our efforts will focus on three objectives. First, every person in America should have easy access to reliable information about their communities, their climate, and how these are changing. Second, every business should have easy access to reliable information on its market, potential markets, scientific information, and changing economic conditions. Further, new data-based businesses should be able to easily pull our data, combine it with other information, and make new products to compete in the private marketplace. Third, and finally, every government should have easy access to the information it needs to better serve its communities and to assess the efficacy of its programs.

The following report, *Fostering Innovation, Creating Jobs, Driving Better Decisions: The Value of Government Data*, provides a helpful lay of the land in assessing where we are in our efforts. Focusing in particular on the value of statistical data produced by Federal agencies, this report explains what Government brings to the table in providing data and describes the wide range of decisions that are supported by Government data. Finally, it highlights the importance of collaboration between Government and the private sector by examining the size of the private sector that intensively uses Government data.

As the Government continues to push forward in its efforts to unlock data, this data-driven sector of the economy will continue to grow, guiding decisions across the country, helping make government smarter, businesses more competitive, and citizens more informed.



**Penny Pritzker**  
U.S. Secretary of Commerce

# Fostering Innovation, Creating Jobs, Driving Better Decisions: *The Value of Government Data*

## Acknowledgments

This report was prepared by Regina Powers and David Beede in the Office of the Chief Economist, Economics and Statistics Administration.

The authors would like to thank the following persons who provided comments, suggestions, and other contributions to this report. Of course, any errors in the report are solely the authors' responsibility.

### **Economics and Statistics Administration:**

- Mark Doms, Under Secretary of Commerce for Economic Affairs
- Kenneth Arnold, Deputy Under Secretary of Commerce for Economic Affairs
- Sue Helper, the Chief Economist of the Department of Commerce
- Austin Durrer, Chief of Staff
- Robert Rubinovitz, the Deputy Chief Economist of the Department of Commerce
- Sabrina Montes, Section Chief in the Office of the Chief Economist
- David Langdon, David Henry, Alis Asadurian, Sandra Cooke-Hull, Albert Espinoza Cassandra Ingram, Jessica Nicholson, Ryan Noonan, and Augustus Urschel of the Office of the Chief Economist
- Richard Lattimer and Kenneth White of the Office of Administration
- Lucas Hitt, Acting Director for Communications (on detail from the Bureau of Economic Analysis), Jane Callen, Editor, and Jacque Mason, Communications Liaison

**Office of Management and Budget:** Paul Bugg, Melissa Chiu, and Brian Harris-Kojetin

**Bureau of Labor Statistics:** Erica L. Groshen, Commissioner; John Eltinge, Associate Commissioner for Survey Methods Research; John W. Ruser, Associate Commissioner for Productivity and Technology; David Friedman, Assistant Commissioner for Industrial Prices and Price Indexes of the Bureau of Labor Statistics; Michael D. Levi, Associate Commissioner, Office of Publications and Special Studies; and Kathy Frederickson-Mele, Deputy Associate Commissioner, Office of Publications and Special Studies

**Bureau of Economic Analysis:** Brian C. Moyer, Deputy Director and Acting Director; Dennis J. Fixler, Chief Statistician; Kathleen James, Chief Administrative Officer; Brent R. Moulton, Associate Director for National Economics; Sarahelen Thompson, Associate Director for International Economics; Joel D. Platt, Associate Director for Regional Economics; and Carol Moylan, Associate Director for Industry Economics

**U.S. Census Bureau, Economic Programs Division:** Shirin A. Ahmed, Assistant Director for Economic Programs, and Lisa Blumerman (now Assistant Director for Decennial Census Programs), William Davie, Carrie Dennis, Charles Funk, Mendel Gayle (now retired), Andrew Hait, Sam Jones, Steve Owens, Matt Przybocki, and Deborah Stempowski

**U.S. Census Bureau, Demographic Division:** Enrique Lamas, Associate Director for Demographic Programs, Eloise Parker, Assistant Director for Demographic Programs, and David Johnson (now Chief Economist, Bureau of Economic Analysis), Ruth Ann Killion, Timothy P. Olson, Barry Sessamen, Victoria Velkoff, and David Waddington

**U.S. Census Bureau, Research and Methodology Division:** Thomas A. Louis, Associate Director for Research and Methodology and Chief Economist; Ron S. Jarmin, Assistant Director for Research and Methodology; and Lucia S. Foster, Chief Economist and Chief of the Center for Economic Studies

# Fostering Innovation, Creating Jobs, Driving Better Decisions: *The Value of Government Data*

## Executive Summary

Everyone is talking about the importance of data to our society as data improves all of our decisions: those we make as individuals, as businesses, as governments.

Government has been in the data business for quite some time, going back to the first Decennial Census in 1790. Since then, the U.S. Government has played a key role in providing valuable data to our country.

Just how valuable is the data the Federal Government provides? That question can't be answered precisely, but there are many reasons to believe that the value of the data to our society far exceeds its cost. As a first step in ascertaining better estimates of just how valuable Government data is, this report focuses on Federal statistical data, data that informs us about our huge, complex, and dynamic economy and data that tells us about our ever changing population. Examples of Government statistical data include gross domestic product, employment and unemployment, consumer prices, retail sales, housing vacancies, residential construction, agricultural supply and demand, corporate profits, and international trade; there are many more.

### This report finds:

#### **Government data potentially guides trillions of dollars of investments each year.**

Government data helps governments to better target scarce resources, businesses stay competitive, and individuals stay informed about the communities in which they live. As the real world examples in this report demonstrate, individuals, businesses, other organizations, and governments use Government data to help make better informed decisions that are better, faster, and more plentiful because of the ready availability and high quality of Government data.

#### **The cost of Government data is small relative to its potential benefits.**

Since 2004, the Federal Government's principal statistical agencies have spent an average of \$3.7 billion annually on data collection, processing, and dissemination. This expenditure amounts to about three cents, per person, per day, and is only 0.02 percent of our roughly \$17 trillion dollar economy.

#### **Government data is uniquely comprehensive, consistent, confidential, credible, relevant, and accessible.**

Acting alone, the private sector would likely provide only some of the types of data produced by the Government. The Federal Government is uniquely positioned to provide comprehensive, consistent, credible, relevant and accessible data, all while protecting confidentiality. Government data is used directly to support decision making, indirectly through commercially available value-added products, and as a benchmark and standard for private data products. Thus, the Government and the private sector complement each other.

## **Government data is commercially valuable.**

Government data is a key input to a wide variety of commercial products and services in the economy, although many of these uses may not be apparent because attribution to the Government is not required. This report identifies industries that use Government data intensively and provides a rough estimate of the size of this sector. The lower-bound estimate, based on a very short and incomplete list of firms that rely heavily on Government data, suggests that Government data helps private firms generate revenues of at least \$24 billion annually—far in excess of spending on Government statistical data. The upper-bound estimate suggests that this sector generates annual revenues of \$221 billion. These crude estimates provide rough order-of-magnitude estimates of the range of the sector's size and illustrate the importance of Government data as an input into commercial products and services.

The value of Government data is the extent to which the decisions based on Government data are better than the decisions that would have been made without Government data. Because such counter-factual evidence is not widely available, this report makes the case that the value of Government data far outweighs its cost by outlining the economic arguments for Government provision of data, by estimating the size of the industry sector that intensively uses Government data, and by illustrating the wide range of ways in which Government data is used.



# Table of Contents

Foreword .....	1
Acknowledgments .....	2
Executive Summary .....	3
Table of Contents .....	5
I. Introduction .....	7
II. Government Data: Statistics for Better Decisions .....	9
What is Government Data? .....	9
Spending on Government Statistics .....	11
Table 1: Principal Statistical Agencies of the Federal Government .....	11
Figure 1: Federal Government Spending on Principal Statistical Agencies .....	12
Figure 2: Federal Budget for Selected Functions, 2004-2013 Average .....	13
Table 2: Gross Domestic Investment .....	14
Government Data Enables Better Decision Making .....	14
III. The Role of Government in Providing Data .....	19
Addressing Market Failures to Provide More Information .....	19
The Advantages of Government Production of Statistics and Other Data .....	22
The Complementary Relationship between Government and Private Sector Data .....	29
IV. The Commercial Value of Government Data .....	31
Defining Government Data-Intensive Business Activities .....	31
Figure 3: Government Data Value Chain .....	32
Estimating the Size of the Government Data-Intensive Sector .....	33
Table 3: Revenues and Employment of Firms in the Government Data-Intensive Sector .....	40
Table 4: Revenues of Government Data-Intensive Sector (GDIS) by Industry .....	41
V. Conclusion .....	43
References .....	45
Appendix I: Definitions of Terms Used .....	51
Appendix II: Definitions of Government Data-Intensive Industries .....	53



# I. Introduction

Modern advances in collecting, processing, disseminating, and preserving information have resulted in the proliferation of data from a wide variety of sources. The increase in the amount and variety of data that is available has highlighted long-standing questions about the role of the Federal Government in collecting and disseminating data. Why does the Government, in certain cases, mandate the collection of personal and private data from individuals, households, and businesses? Are collecting and disseminating data appropriate activities for Government or better left to the private sector? What happens to the data after the Government collects it? Do the benefits of the data outweigh the costs?

It is fairly straightforward to empirically estimate the costs to the Government of collecting, processing, and disseminating data. Measuring the benefits of Government data—the value created when people, businesses, and institutions use the data to make better decisions—is more challenging. Unlike goods and services sold on the open market, Government data—like most services provided by the Federal Government—is freely available to the public. Because Government data is not traded in the marketplace, it has no price that can be used to estimate its value.

Generally, people are unaware of the full range of both public and private sector products and services for which Government data is a vital ingredient. One reason that many uses of Government data are not visible is that there are no copyrights for Federal Government data and anyone can use the data without attribution. The combination of free dissemination of Government data and its use without required attribution has made it difficult to identify all users and uses of the data, further complicating efforts to estimate the value of the data. Because we cannot observe how decisions would be made in the absence of Government data, it is difficult to estimate the incremental value of the improvement in decision making that is made possible with Government data. Despite the measurement challenges, it is worthwhile to understand the value of Government data, particularly given the rapidly changing landscape of data production and dissemination.

The remainder of this report is organized as follows:

Section II defines Government data and discusses the costs of producing Federal statistics, as well as how those statistics are used for decision making that affects people throughout the Nation. In addition, sidebars throughout the report provide examples of data-driven decisions made by individuals, businesses, other organizations, and governments informed by Government data.

Section III discusses the rationale for Government provision of data and, focusing on the Federal statistical system, describes why the Government is involved in providing data and outlines the advantages Government often has over the private sector in collecting and disseminating statistical data. It also discusses the important role of the private sector in complementing and adding value to Government data to meet specific information needs in the market.

Section IV further explores the commercial value of Government data, illustrating the extent of the use of Government data by the private sector. It discusses estimates of the size of the Government data-intensive sector (GDIS)—the private industries that rely on Government data as a critical input to the goods and services they produce.

Section V summarizes the limited evidence about the value of Government data and concludes that the commercial value of Government data, as approximated by the revenues of data-intensive firms, far exceeds the costs of Government data. Moreover, because of its high quality and public good characteristics, the total value of Government data is greater than just its commercial value.



## II. Government Data: Statistics for Better Decisions

### What is Government Data?

The term “Government data” includes diverse categories of Government information products, such as the following:

- *Information published by Government agencies for statistical purposes.* Such information is compiled from raw data collected directly from individuals, households, businesses, other organizations, and governments. The information may also be indirectly compiled from Government administrative data,<sup>1</sup> private sector data, and other sources. Statistical agencies summarize the data in tabulations to describe specific characteristics of groups of persons, businesses, or other entities. The agencies make such information available for use by other Government agencies and the general public to inform decision making, in program analysis and evaluation, and in research.
- *Information in Government administrative records on individuals and businesses.* Examples include publicly available data about private sector company-sponsored retirement plans reported to the Labor Department on Form 5500; Securities and Exchange Commission filings; Government contracts data; real estate appraisal records; and voter registration rolls. Historically, a Federal or State Freedom of Information Act request has often been required to obtain access to this category of information, but recent “open data” efforts have aimed to make such information more easily accessible. Other data is strictly confidential and therefore not publicly released, such as tax return data collected by the Internal Revenue Service or Social Security Administration data.
- *Physical measurements about natural phenomena recorded by Government agencies.* Examples include weather and climate data such as local temperature readings collected by the National Weather Service of the National Oceanic and Atmospheric Administration, Global Positioning System data, and measurements of the characteristics of elements and molecules collected and published by the National Institute for Standards and Technology Physical Measurement Laboratory.

#### Box 1: Data versus Statistics versus Information

The terms “data,” “statistics,” and “information” each have distinctly different formal definitions. Data is the least abstract concept—it consists of a list of discernible variations in the characteristics of measurable states-of-the-world (Boisot and Canals, 2004). For example, a datum may be the fact that person A lives in New York City, is married, and earns \$73,000 per year; while another datum may be the fact that person B lives in Omaha, is unmarried, and earns \$42,000 a year. Statistics are measures of some attribute of a set of data, such as the average earnings per year of unmarried persons living in New York City. Information is a set of noteworthy regularities embedded in the data that can contribute to knowledge (Boisot and Canals, 2004). In common parlance, the distinctions between these concepts are blurry, and they are often used interchangeably, including in this report. The term data is commonly used to convey the notion of a body of information, as well as to convey the notion of uniquely identifiable facts, giving rise to vigorous debate over whether the term is singular or plural. To capture the more general understanding of data as information, this report treats the term data as a singular noun.

<sup>1</sup> Administrative data is data collected and used by program, administrative, and regulatory agencies and not originally intended to be used for statistical purposes. See Appendix I for definitions of other terms used in this report.

In this report, we focus on a subset of the first category: Government data published by the Principal Federal Statistical Agencies within the Executive Branch of the Federal Government. These agencies' core missions are to collect, compile, process, analyze, and disseminate information for statistical purposes. Table 1 lists these agencies and their average annual budgets for 2004-2013. It is notable that these budgets do not account for all spending on Federal statistics. In fact, funding for the statistical programs of the Principal Statistical Agencies (excluding the Decennial Census, which is associated with a surge in expenditures every decade) accounts for about one-third of all Government funding for statistics (Office of Management and Budget, 2013a, p. 4). The remaining two-thirds share of the funding for statistical purposes is devoted to programs conducted within various agencies to help them carry out their missions. For example, the Environmental Protection Agency collects data to enforce environmental regulations; the Occupational Safety and Health Administration collects data to enforce occupational health and safety regulations; and the Employee Benefits Security Administration collects data to enforce regulations of employee retirement benefits (Office of Management and Budget, 2013a, p. 5).

### **Box 2: BrightScope Creates Value-Added Products Using Data Embedded in Public Documents**

BrightScope is a financial information company that provides ratings and investment analysis of retirement plans for participants and sponsors, asset managers, and financial advisors. The company maintains a comprehensive database of information on the retirement plan market and produces quantitative ratings of over 45,000 401(k) and 403(b) plans. These ratings are based on more than 200 unique data inputs per plan. BrightScope also publishes distribution rankings for funds in the retirement marketplace, allowing comparisons of the performance of asset managers. The company markets a suite of data analytics software products to Fortune 1000 companies, asset managers, broker-dealers, financial advisors, and other market participants. Definitions, criteria, and methodologies are available for free at the company's website. BrightScope also encourages and enables plan sponsors and participants to upload current data about their retirement plans.

([www.brightscope.com/about/](http://www.brightscope.com/about/))

Brightscope relies upon key information from the Department of Labor's Form 5500 Series. Employee benefit plans use the Form 5500 Series to satisfy annual reporting requirements under Title I and Title IV of the Employee Retirement Income Security Act of 1974 (ERISA) and under the Internal Revenue Code. The Form 5500 is an important compliance, research, and disclosure tool for the Department of Labor, a disclosure document for plan participants and beneficiaries, and a source of information and data for use by other Federal agencies, Congress, and the private sector in assessing employee benefit, tax, and economic trends and policies. The Form 5500 Series is part of ERISA's overall reporting and disclosure framework, which is intended to assure that employee benefit plans are operated and managed in accordance with certain prescribed standards and that participants and beneficiaries, as well as regulators, are provided or have access to sufficient information to protect the rights and benefits of participants and beneficiaries under employee benefit plans.

(<http://www.dol.gov/ebsa/5500main.html>)

## Spending on Government Statistics

On average between 2004 and 2013, the United States spent \$3.7 billion annually, adjusted for inflation, on data collection and dissemination by the Principal Statistical Agencies (Table 1).<sup>2</sup> The Decennial Census is associated with a surge in expense. Excluding the Decennial Census, the average is about \$2.3 billion. (See Figure 1.)

**Table 1: Principal Statistical Agencies of the Federal Government**

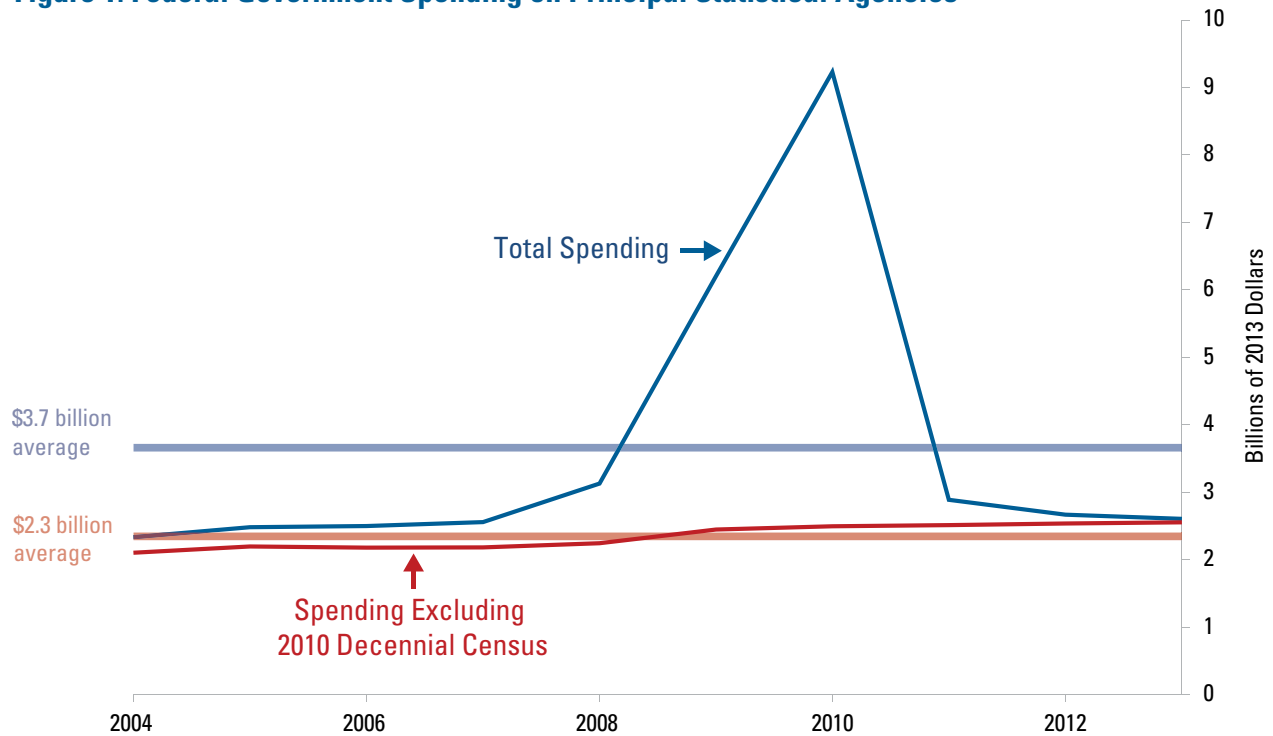
Agency	Parent Department or Agency	Average Annual Budget Authority 2004-2013 (millions of 2013 dollars)
Bureau of Economic Analysis	Commerce	\$90
Bureau of Justice Statistics	Justice	55
Bureau of Labor Statistics	Labor	620
Bureau of Transportation Statistics	Transportation	29
U.S. Census Bureau	Commerce	2,029
Economic Research Service	Agriculture	84
Energy Information Administration	Energy	104
National Agricultural Statistics Service	Agriculture	162
National Center for Education Statistics	Education	244
National Center for Health Statistics	Health and Human Services	135
National Center for Science and Engineering Statistics*	National Science Foundation	35
Office of Research and Statistics	Social Security Administration	28
Statistics of Income Division	Treasury	42
<b>Total</b>		<b>\$3,657</b>

*Note:* Budget information compiled from *Statistical Programs of the U.S. Government* and actual agency budgets. Budget amounts include spending on the Decennial Census and are converted to 2013 dollars using the Federal Government Consumption Expenditures price index.

\* Formerly the Science Resources Statistics Division.

<sup>2</sup> Another component of the cost of statistical data is the opportunity cost of the time spent by persons filling out survey forms. We do not address that cost in this report, although we note that The Paperwork Reduction Act (44 U.S.C. §§ 3501-3521) is intended to minimize the burden on respondents to Federal surveys in order to “ensure the greatest possible public benefit from and maximize the utility of information created, collected, maintained, used, shared and disseminated by or for the Federal Government.”

**Figure 1: Federal Government Spending on Principal Statistical Agencies**

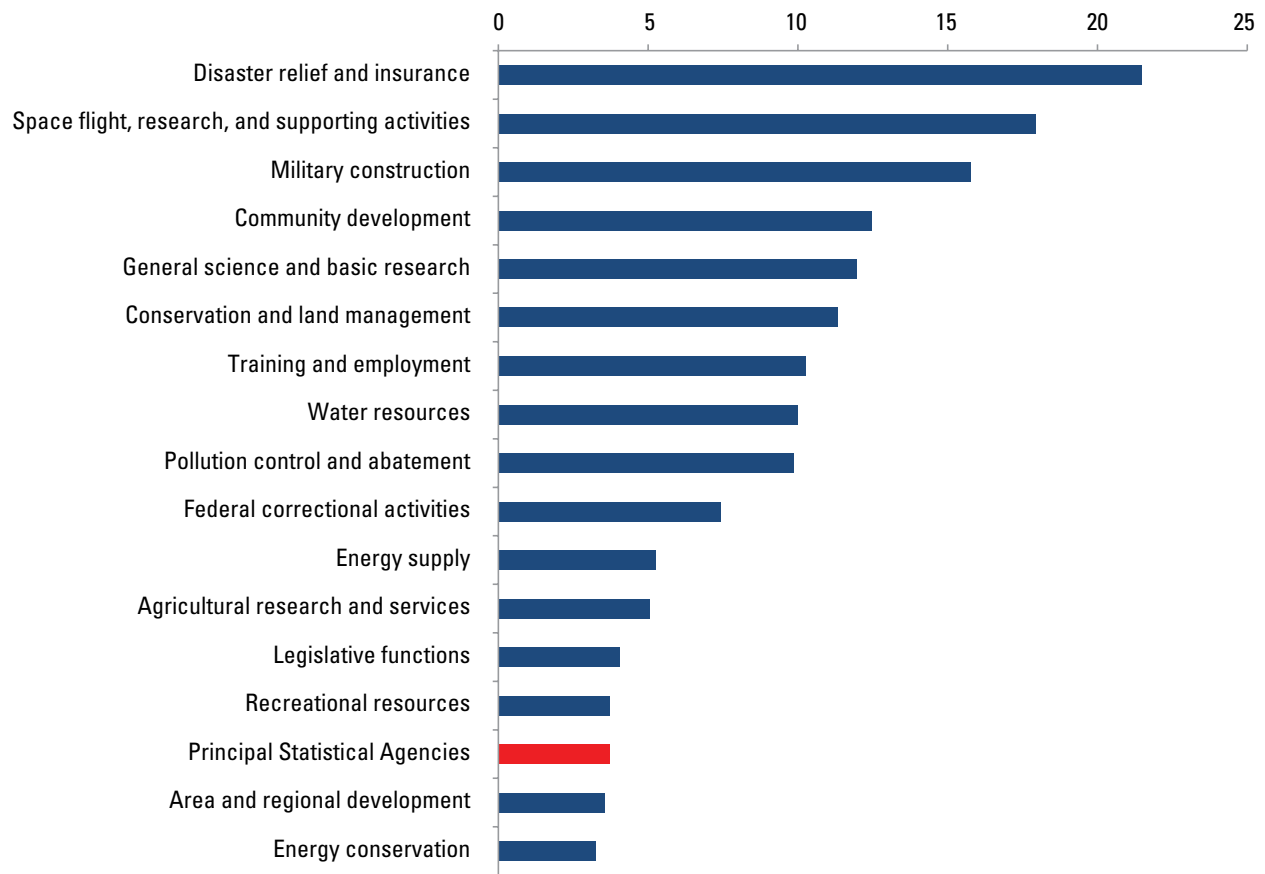


Sources: Budget information compiled from *Analytical Perspectives, President's Budget, Statistical Programs of the U.S. Government Supplement to President's Budget*; actual agency budgets; *Principles and Practices for a Federal Statistical Agency*  
 Note: Budget amounts converted to real 2013 dollars using Government Consumption Expenditures deflator.

Figure 2 puts the \$3.7 billion average annual spending by the Principal Federal Statistical Agencies in perspective by comparing it to spending on selected other functions of the Federal Government. This chart illustrates that spending on the Principal Statistical Agencies is relatively small compared to spending on many other important government functions. Another way to put this spending in perspective is to realize that \$3.7 billion is roughly 0.022 percent of 2013 U.S. Gross Domestic Product of nearly \$17 trillion, and only \$11.57 per capita per year. That's about equal to three cents, per person, per day.



**Figure 2: Federal Budget for Selected Functions, 2004-2013 Average**  
(Billions of 2013 Dollars)



Source: Office of Management and Budget (2013b). For Principal Statistical Agency total, see Table 1.

Government data also could be considered a form of investment, even though they are not currently treated as investments in the National Income and Product Accounts.<sup>3</sup> Employment, Gross Domestic Product, and other Federal Government statistics are information assets built up from cumulative spending on data collection, processing and analysis, much like knowledge accumulated from years of spending on research and development. Therefore, yet another way to put expenditures on Government data in perspective is to note that the \$3.7 billion average annual expenditure on (or investment in) the Federal Principal Statistical Agencies is approximately one-tenth of one percent of total annual economy-wide investments, that is, 0.11 percent of the \$3.276 trillion average annual total gross domestic investment for 2013 (see Table 2).

<sup>3</sup> In contrast, expenditures on research and development are treated as investments rather than expenditures in the National Income and Product Accounts.

**Table 2: Gross Domestic Investment**  
(Billions of 2013 Dollars)

Domestic business	\$2,108
Households and institutions	562
Federal Government	275
State and local governments	332
<b>Total</b>	<b>\$3,276</b>

Source: Bureau of Economic Analysis, 2014.

## Government Data Enables Better Decision Making

Information is valuable to the extent it affects action (Hirshleifer, 1972, p. 564). Information produced by the Federal Government affects action by aiding people in their decisions. Government data thus increases in value when it improves people's decisions. This report illustrates how Government data informs investment decisions by people, businesses, other organizations, and governments at every level, providing evidence that, absent Government data, the loss of information would degrade the quality of decision making. In particular, given the pervasive use of Government statistics to inform investment decisions throughout the economy, the absence of such data would likely result in an annual reduction in the return on aggregate investment far greater than the \$3.7 billion average annual spending associated with producing the data.

Government data provides important fact-based evidence to support a wide range of decisions in the public and private sectors. The data used for these decisions may come directly from a Government agency through its public dissemination activities, or it may come indirectly from third party intermediaries. In any event, one important concept in thinking about the value of Government data lies in the extent to which these evidence-based decisions are better

### **Data-Driven Decisions: Improving Downtown Life**

*The Center City District (CCD) is a Business Improvement District in downtown Philadelphia. This organization was formed to keep Philadelphia's downtown clean, safe, beautiful, and fun. CCD makes physical improvements to the downtown, such as installing and maintaining lighting, signs, banners, trees and landscape elements (Center City District, 2014). This organization used data from the Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program on where people live and work to better define its downtown, thereby improving the focus of its \$20 million operating budget. The LEHD program combines data from a wide range of sources, including Unemployment Insurance earnings data from States, data from the Quarterly Census of Employment and Wages, and data from the Decennial and Economic Census surveys, to create statistics on employment, earnings, and job flows at detailed levels of geography and industry and for different demographic groups. The LEHD also generates partially synthetic data on workers' residential patterns (Gilchrest, 2013).*

*The pool of users of Government data for local economic development and other local public policy making is huge; Census Bureau data shows that there are 1,249 principal cities in metropolitan and micropolitan statistical areas in the United States; 729 incorporated places with a population of 50,000 or more; and over 3,140 counties. In addition, there are over 40,000 members of the American Planning Association (American Planning Association, 2014). In fact, non-profit groups re-package and tailor Government data specifically to analyze local issues. For example, the California Business Roundtable has set up a non-profit organization called the Center for Jobs and the Economy that will provide a free online and interactive repository of Federal data from the Census Bureau and the Bureau of Labor Statistics, as well as California state agencies such as the Employment Development Department (Walter, 2013).*

than the “best guess” decisions that would have been made without the data. Assessing value in such a manner would require a body of counterfactual evidence that does not currently exist. In the absence of such counterfactual evidence, we focus instead on identifying and describing the wide range of decisions supported by Federal Government data.

### **Box 3: Government Data from the National Oceanic and Atmospheric Administration Parallels with the Principal Federal Statistical Agencies**

As noted above, the National Oceanic and Atmospheric Administration (NOAA) of the U.S. Department of Commerce collects weather and climate data. Research examining that value suggests that the return to society on investment from Government meteorological data is large. For example, one study found that the overwhelming majority of survey respondents reported using weather forecasts and did so at an average rate of 3.8 per day (Lazo, et al., 2009).<sup>\*</sup> This means that, based on Census Bureau population estimates, U.S. adults obtained about 301 billion forecasts in the year the study was conducted. The researchers also found a median valuation of weather forecasts per household of \$286 per year, which suggests that the aggregate annual valuation of weather forecasts was about \$31.5 billion. (Lazo, et al., 2009) The sum of all Federal spending on meteorological operations and research was \$3.4 billion in the same year, and the private sector spent an additional \$1.7 billion on weather forecasting, for a total of private and public spending of about \$5.1 billion. In other words, the valuation people placed on the weather forecasts they consumed was 6.2 times as high as total expenditure on producing forecasts.

NOAA data is re-packaged and analyzed (that is, made suitable and available for mass consumption by private sector businesses) to produce 15 million weather products (Konkel, 2013). Often the entities that use data produced by NOAA’s National Weather Service (NWS), such as The Weather Channel, AccuWeather, and numerous other private sector forecasters, do so without attributing NOAA as the source of the data. As a consequence, many end-users may not realize the value that NOAA’s data provides. Private sector firms continually develop new ways of creating value for companies using data from NOAA and other Government agencies. For example, the Climate Corporation sells weather insurance to farmers and is also a high-tech agricultural consultant. The company processes 50 terabytes of new and historical data each day. One of its principal data sources is the National Weather Service’s Nexrad (Next Generation Radar), a network of 159 Doppler radar stations that scans weather data in two million locations. Another source of data is 60 years of crop-yield statistics from the Department of Agriculture. Together, these data sources enable the company to generate moisture, precipitation, and soil condition maps down to the level of a farmer’s field. Monsanto bought the company in November 2013 for nearly one billion dollars (Spector 2013).

The private sector adds value to only two of the 19 terabytes of new weather and climate data that the agency produces every day; the rest currently is not accessible to the public. NOAA is working to make the data more accessible. NOAA is developing public-private partnerships in which private companies host NOAA data on the cloud and make it readily available to the public for advanced analytics without having to download all the data (Konkel 2013).

<sup>\*</sup>Incidentally, the researchers used the Census Bureau’s American Community Survey as a benchmark to determine that the respondents were generally representative of the U.S. population—an example of using Government data for benchmarking purposes to help validate statistical results from other surveys.

**Data-Driven Decisions: Screening for Cancer**

*The Centers for Disease Control and Prevention (CDC) spent \$150 million in 2013 on cancer screening services and outreach, with the potential for saving lives through early detection in tens of thousands of women. CDC's National Breast and Cervical Cancer Early Detection Program, as well as State-based programs, need to determine where to target outreach and screening services based on the percentage of women eligible for early detection services. In order to achieve this objective, CDC helps sponsor the Census Bureau's Small Area Health Insurance Estimates (SAHIE). These are model-based estimates that combine American Community Survey data along with data from the Census Bureau's demographic population estimates, 2010 Census, County Business Patterns; the Internal Revenue Service's Federal tax information; the Food and Nutrition Service's Nutrition Assistance Program; and the Centers for Medicare and Medicaid Services' (CMS) Children's Health Insurance Program and Medicaid. In addition to informing decisions on where to target cancer screening efforts, SAHIE provides yearly estimates of health insurance coverage for every county in the United States. SAHIE's annual operating cost of \$2 million thus helps CDC and other agencies to make better decisions on spending much larger sums of money on preventative health services than would be the case absent such data.*

From the beginning of our Nation's history, there was recognition of the need for data in order to make informed decisions. Collecting data was an essential activity envisioned by the country's founding fathers, and the taking of a census at regular intervals for the apportionment of Representatives in Congress was written into the U.S. Constitution in 1787.<sup>4</sup> The first Decennial Census was taken in 1790; its purpose was to ensure a fair and accurate system of representation and taxation.

Over time, as the country's economy developed, the need for accurate information grew beyond a simple population count. Collection of additional information built on the foundation set by the early censuses. In the early 1800s, Congress directed that a census of manufactures be conducted in order to provide a "general view of the manufactures of the United States." This census was followed by a census of agriculture and mineral industries (Duncan and Shelton, 1978). By the late 1800s, the Federal Government was publishing statistics on foreign trade,

government expenditures and the internal commerce in the country. As the U.S. economy and society grew and changed in the 20th Century, the need for statistical information increased still further.

Today, data from Federal statistical agencies informs a multitude of decisions that help government improve its operations. One important function of Government data is to help direct the funds to programs and areas where they are most needed, such as the National School Lunch program. For example, in 2012, 28 Federal Government programs used regional income and product estimates from the Bureau of Economic Analysis to distribute \$320 billion; 27 Federal programs used Local Area Unemployment Statistics from the Bureau of Labor Statistics to distribute \$115 billion, and the Census Bureau's American Community Survey was used to help determine how over \$400 billion in Federal and State funds were distributed.

Government data also helps Government agencies to ensure that the benefits of their regulations justify their costs. For example, in 2013, Federal agencies evaluated the impact of over 100 "economically significant" regulations, each with a potential impact of at least \$100 million or otherwise materially affecting the economy (Office of Information and Regulatory Affairs 2014). Without statistical and administrative data on the number of firms affected and the revenues of those firms, for example, estimates of the benefits and costs of these regulations would be more uncertain. As a result, there would be a greater risk that costly regulations would be enacted, while beneficial regulations would not.

<sup>4</sup> Article I, Section 2, Clause 3 of the U.S. Constitution states, "Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers . . . . The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such Manner as they shall by Law direct. . . ." This Clause was amended in 1868 by Section 2 of the 14th Amendment to provide that apportionment shall be based on "counting the whole number of persons in each State, excluding Indians not taxed."

Government statistics, most notably the Principal Federal Economic Indicators, are carefully watched by decision makers across the globe in order to track the macro-economy. These decision makers include Government officials at the highest levels, who use the data to gauge the overall economic health of the Nation, prepare the Federal budget, and project tax revenue; the data also informs monetary and trade policies.

#### **Box 4: The Principal Federal Economic Indicators**

The Principal Federal Economic Indicators are statistical series designated by the White House Office of Management and Budget that provide timely measures of economic activity. These indicators are published by several statistical agencies and are listed below. Because these series can have important impacts on market decisions and government policies, they are evaluated every three years, including an analysis of their accuracy and performance and an assessment of the effects of revisions.

##### **DEPARTMENT OF AGRICULTURE, WORLD AGRICULTURAL OUTLOOK BOARD**

- World Agricultural Supply and Demand Estimates

##### **DEPARTMENT OF AGRICULTURE, NATIONAL AGRICULTURAL STATISTICS SERVICES**

- Agricultural Prices
- Crop Production
- Grain Stocks
- Cattle on Feed
- Hogs and Pigs
- Plantings

##### **DEPARTMENT OF AGRICULTURE, FOREGIN AGRICULTURAL SERVICE**

- World Agricultural Production

##### **DEPARTMENT OF COMMERCE, U.S. CENSUS BUREAU**

- Construction Put in Place
- New Residential Construction
- New Residential Sales
- Monthly Wholesale Trade
- Advance Monthly Sales for Retail and Food Services
- U.S. International Trade in Goods and Services
- Manufacturing and Trade: Inventories and Sales
- Manufacturers' Shipments, Inventories, and Orders
- Advance Report on Durable Goods—Manufacturers' Shipments, Inventories, and Orders
- Quarterly Financial Report—Manufacturing, Mining, and Wholesale Trade
- Quarterly Financial Report—Retail Trade
- Housing Vacancies
- Quarterly Services

##### **DEPARTMENT OF COMMERCE, BUREAU OF ECONOMIC ANALYSIS**

- Personal Income and Outlays
- Gross Domestic Product
- Corporate Profits
- U.S. International Trade in Goods and Services
- U.S. International Transactions

##### **DEPARTMENT OF ENERGY, ENERGY INFORMATION ADMINISTRATION**

- Weekly Natural Gas Storage Report

##### **DEPARTMENT OF LABOR, BUREAU OF LABOR STATISTICS**

- The Employment Situation
- Producer Price Indexes
- Consumer Price Index
- Real Earnings
- Productivity and Costs
- Employment Cost Index
- U.S. Import and Export Price Indexes

##### **FEDERAL RESERVE BOARD**

- Money Stock Measures
- Factors Affecting Reserve Balances of Depository Institutions and Condition Statement of Federal Reserve Banks
- Industrial Production and Capacity Utilization
- Consumer Installment Credit

It is noteworthy that decision makers often rely on data originating from a variety of Government agencies, and even many statistical agencies rely on the data products of other agencies. For example, the Bureau of Economic Analysis (BEA) estimates of national accounts, including Gross Domestic Product, depend critically on data from the Economic Census produced by the Census Bureau. Similarly, the Census Bureau's model-based Small Area Income and Poverty Estimates (SAIPE) rely on data from various Census Bureau programs, as well as data from the Internal Revenue Service's Federal tax information, the Social Security Administration's Supplemental Security Income, the BEA's Regional Economic Income Statistics, and the Food and Nutrition Service's Supplemental Nutrition Assistance Program. The SAIPE program serves as the only source for yearly estimates of school-aged children in poverty for all States, counties, and school districts.<sup>5</sup>

### ***Data-Driven Decisions: Transportation Planning***

*In 2013, the value of construction put in place for highways and streets averaged \$81 billion—and the decisions leading to that construction can be informed by census and other Government data (U.S. Census Bureau, 2013b). For example, the San Diego Association of Governments (SANDAG) adopted its 2050 Regional Transportation Plan (RTP) in 2011, with planning recommendations based on models that used data from the American Community Survey, such as population, housing, household structure, active-duty military personnel and their dependents, and school enrollment, to forecast neighborhood populations and to predict future travel patterns (National Research Council, 2013, p. 24-28). Additional data was drawn from Census Transportation Planning Products (CTPP), a Census Bureau program widely used by transportation planners. Many of the capital projects outlined in the transportation plan are currently in development.*

There are many other examples of cross-agency use of data. The Administrative Office of the U.S. Courts uses Decennial Census data on voting age, sex, race, and Hispanic origin to assist district courts in determining whether their jury pools comply with the Jury Selection and Service Act.<sup>6</sup> The Social Security Administration relies on data from the Consumer Price Index published by the Bureau of Labor Statistics to adjust benefit payments for inflation.

The Census Bureau's Equal Employment Opportunity Tabulation is the primary benchmark for comparing the race, ethnicity, and sex composition of an organization's workforce within a specified geography and job category. The Equal Employment Opportunity Commission, the Department of Justice, the Department of Labor, and the Office of Personnel Management use the data to monitor workforce decisions of private sector businesses, Federal contractors, Federal Government agencies, and state and local government agencies.

The government decisions supported by various sources of Government data result in widely diffuse benefits that are difficult to quantify but that nonetheless illustrate the value of the data. Throughout this report, several sidebars provide additional examples of decisions made by individuals, businesses, other organizations, and governments that are informed by Government data. These examples further highlight the value of Government data.

<sup>5</sup> The SAIPE program, with a cost of under \$4 million, results in data used to allocate \$15.4 billion (in 2013) to school districts and States, assisting more than 16 million school-aged children in poverty.

<sup>6</sup> Codified as amended at 28 U.S.C. §§ 1861 – 1878 (2012).

### III. The Role of Government in Providing Data

Making decisions under uncertainty is a fact of life. For example, individuals and households must continually decide on their levels of consumption (including leisure) and savings. Periodically they must make major life decisions such as choosing a career, when and where to search for a job, getting married, starting a family, making major purchases such as homes and cars, investing in education and training, and planning for retirement. Businesses have to make investments and operational decisions about such questions as what to produce, how much to produce, where to produce it, how to compensate employees, and how to protect themselves from price risks. Governments have to decide on the types and quantities of services to provide and where to provide them. Common to all these problems is a need for information to guide the decision maker in choosing the best possible action from among the many possible actions. From an economist's perspective the best possible action is the one that provides the highest overall benefits relative to its costs.

Better information allows people, businesses, other organizations, and governments to make better decisions. For example, detailed data about a neighborhood's characteristics (such as average residence square footage and average number of household members) can help a retailer better understand its consumers and therefore make better decisions about the types of merchandise to carry for sale in that neighborhood. If such data were not available, the retailer would have to either find a way to collect the data—a costly investment—or make do with inferior data that might lead to poorer merchandising decisions. In this section, we discuss how Government provision of data provides decision makers with more and better information than would otherwise be the case.

#### ***Data Driven Decisions: Customizing Retail Product Selections***

*Retail companies go to great efforts to understand their markets and tailor the products they offer. Data about local communities often weighs heavily in analyses of communities.*

*For example, Target Corporation, with more than 1,700 stores in 49 states and employing more than 300,000 persons, uses American Community Survey (ACS) data to tailor its merchandise offerings for customers in its neighborhoods. Compared to suburban residents, persons living in urban areas live in smaller spaces, have fewer members in their households, and are more likely to use public transportation. Based on this information, Target stocks its urban stores with smaller furniture, more folding chairs, and smaller package sizes than its suburban stores. Moreover, Target uses ACS data to identify which urban neighborhoods have attracted young professionals, so it carries apparel in more youthful fashions in those stores than in its stores in areas with older population, such as some of its suburban stores (U.S. Census Bureau, 2012).*

#### **Addressing Market Failures to Provide More Information**

Government data is an information asset built up from cumulative spending on data collection, processing and analysis, with characteristics that distinguish it from products that are more efficiently produced by markets. Information products—including Government data—have special characteristics that raise the likelihood that the market will fail to produce information at a socially optimal level. Below we discuss these characteristics in more detail.

Products that can be produced efficiently by the market share five features.<sup>7</sup>

1. They are “rivalrous,” that is, only one economic agent can consume a given unit of such products. (If Jack eats a slice of bread, Jane cannot eat his slice—she has to pay to get her own slice of bread if she wants one.)
2. They are easily excludable, such that the producer has property rights to compel a consumer to pay for the product. (Thus, Jack cannot legally eat a slice of bread without paying baker Tom.)
3. They have characteristics and benefits that are “open” in the sense that they are easy to recognize or understand; generally, consumers know what they want and whether it is worth the price at which it is being sold. (Jack and Jill understand what they are buying when purchasing a loaf of bread.)
4. They generally have constant returns to scale or possibly decreasing returns to scale, at least above some level of output. (It takes twice the amount of flour, yeast, water, and other ingredients to make a slice of bread for Jack and for Jill than just for Jack alone.)
5. The production and consumption of such products generally do not impose gains or losses on others; that is, they generate no externalities. (All the costs of baking the bread, and revenues from selling it, accrue to baker Tom; and the expenditure to buy the bread, and the utility gained by eating it, are Jack’s alone.)

In contrast, information generally does not share these five characteristics.

1. The consumption of information is “non-rivalrous.” One person’s use of the information does not diminish another’s use of it.
2. Although information can be made to be excludable (for example, by encrypting it and obtaining and enforcing intellectual property rights to it), it is often costly, or against public policy, to do so. These two characteristics of information—non-rivalry and costly excludability—suggest that information has characteristics of a “public good.”<sup>8</sup> Standard economic theory shows that markets generally provide public goods at less than the socially optimal level; as a result, government intervention could help the economy achieve this socially optimal level.
3. Information, as distinguished from goods that can be efficiently produced by markets, often is opaque; that is, its quality can be difficult to ascertain before acquiring the data. Information is often the product of complex data collection and processing, making it difficult and costly for consumers to assess its quality and reliability. This opacity can lead to lower than optimal production and consumption of information.<sup>9</sup>
4. Information likely has increasing returns to scale at all relevant levels of production, because it has high fixed costs of production (costly collection and processing) and negligible variable costs of production (low costs of dissemination). Moreover, information production can have economies of scope (that

<sup>7</sup> See DeLong and Froomkin, 2000 for a discussion of information and rivalry, excludability, and transparency; see Allen 1990, p. 271 for a discussion of production of information; and see Bates 1988 for a discussion of consumption.

<sup>8</sup> Because information is technically excludable, it may be more accurately characterized as a “club good” rather than a pure “public good.” Nevertheless, when exclusion costs are so high that it is not feasible to exclude non-paying consumers, as may be the case with “headline” data and even more detailed data distributed in electronic format, or when exclusion is contrary to public policy, then technical excludability is less relevant, and the case for public provision of the good is stronger (Cornes and Sandler, 1996, p. 400).

<sup>9</sup> For example, the used car market is arguably less efficient than the new car market because of information asymmetries that make it difficult for a prospective purchaser to ascertain the quality of a used car until it is actually purchased (Akerlof 1970).



is, the average cost of production declines as additional product types are produced), because some of the infrastructure of the production process for one type of information product may also be used to produce another information product. Such increasing returns to scale and scope give rise to natural monopoly conditions, with the result that the supply of information, if left to a monopoly market supplier, would be less than the socially optimal level.

5. The production and consumption of information can generate external benefits. For example, as more people, businesses, and governments use information to make better decisions, the information becomes more valuable. The more that information is used, the more likely it is that new and innovative ways of using the data that improve decision making would be added to the store of general knowledge about information and its beneficial uses.<sup>10</sup> Moreover, users of information may find problems with the information and provide useful feedback to the information producer, or may suggest additional types of information that may enhance the quality and utility of existing data products.<sup>11</sup> In addition, as noted below, setting standards for data, such as definitions of key concepts, makes all data that employs such standards potentially more valuable. In this way, there are network externalities in the production and use of information. Given these externalities, private consumers and providers do not consider the full set of benefits when deciding how much information to exchange. This means that less than the socially optimal amount of information would be produced.

#### ***Data-Driven Decisions: Helping Businesses Export***

*In 2007, a coffee producer in Texas received customized market research from the U.S. Commercial Service office in San Antonio, which helped the company break into markets in Mexico, Italy, Iraq, Australia, and Japan. Market research produced by the Commercial Service includes Government data from Trade Stats Express (<http://tse.export.gov/TSE/TSEhome.aspx>), and even more detailed data is available at the Census Bureau's USA Trade Online (<https://usatrade.census.gov/>).*

*The share of the company's annual income from exports increased from 2% in 2005 to 60% in 2009, and the company added jobs as a consequence. Based on the company's success and knowledge gained from exporting, the owner started an export management company to help other companies (Export.gov, 2014).*

Together, these five characteristics of information—non-rivalry, costly excludability, opacity, economies of scale and scope, and the presence of positive externalities—imply the market would likely produce less than the socially optimal level of information. As a result, there is a role for government intervention to help the economy achieve a more optimal level of information.

<sup>10</sup> Interspersed throughout this report are numerous examples of publicly available descriptions of ways that Government data can be used to improve decision making. These examples, and many more not cited in this report, add to the general store of knowledge about how to use Government data in innovative ways to improve decision making.

<sup>11</sup> For example, Government data agencies “have reported that intensified ties with re-users [of data] may lead to improved data quality and process efficiency since any deficiencies in the data are promptly flagged up and reported back to the” agency. “Hence, when the interest in data quality is shared, quality control is partly outsourced” (POPSIS 2011 p. 6). In addition, there are organizations such as advisory committees established by statistical agencies to obtain advice about better ways to use and improve their data. For example, the Census Bureau and BEA each have advisory committees. The Council of Professional Associations on Federal Statistics (COPAFS) “represents over 300,000 individual researchers, educators, public health professionals, civic groups, and businesses that rely on the quality and accessibility of statistics that can only be effectively collected by the Federal Government” ([www.copafs.com](http://www.copafs.com)). The Federal Committee on Statistical Methodology (FCSM) “is sponsored by the Office of Management and Budget to improve the quality of Federal statistics and the efficiency and effectiveness of statistical practice among Federal agencies” ([www.fcsm.gov](http://www.fcsm.gov)). The Association of Public Data Users (APDU) “is a national network that links users, producers, and disseminators of Government statistical data. APDU members share a vital concern about the collection, dissemination, preservation, and interpretation of public data. As an association, APDU is committed to helping data users identify public data that meet their needs, utilizing data in cost-effective and appropriate ways, establishing effective two-way communication between data producers and users, and bringing the perspectives and concerns of public data users to bear on important issues of Government information and statistical policy. APDU provides a venue to share news, raise concerns, and advocate on behalf of the interests of public data users” ([www.apdu.com](http://www.apdu.com)). Finally, vast numbers of researchers publish peer-reviewed studies using Government data that both lead to improvements in the quality and types of data collected as well as the development of new, useful techniques for using the data.

## The Advantages of the Government Production of Statistics and Other Data

The previous section laid out the reasons that government intervention may lead to a more socially optimal level of information. This section focuses on statistical information in particular and explains the advantages of having the Government produce certain types of statistics and data. The Federal Government is uniquely positioned to provide such data because it provides a system of collection and dissemination that is comprehensive, consistent, confidential, credible, relevant, and accessible.

Federal Government data includes information collected by Federal agencies from individuals, households, businesses, and institutions such as schools, hospitals, and State, local, and tribal governments to serve a wide variety of purposes. The output of Federal statistical programs is one subset of Government data. The statistical data is meant to illuminate public and private decisions on a range of topics, including the economy, the population, the environment, agriculture, crime, education, energy, health, science, and transportation (Office of Management and Budget, 2013a). For example, the Principal Federal Economic Indicators can move markets and shape policy. Government officials at the highest levels use them to gauge the overall economic health of the Nation, prepare the Federal budget, and project tax revenue. They inform fiscal, monetary and trade policies, exchange rates, and social welfare policies, as well as investment and financial decisions of businesses and individuals. Given their importance, there is a compelling need for timely, credible information that is accurate, representative, and comparable.

### 1. Comprehensive

The mission of the Federal statistical agencies is to collect and disseminate nationally representative data that covers the entire population and the whole economy. To ensure accurate representativeness, these agencies collect data from across the U.S. population. Typically this includes population segments—such as persons from sparsely populated regions, or new small businesses—about which private sector entities have little interest. Any single business has little incentive to undertake the significant expense of collecting data from hard to reach populations for their business needs. However, the Government—because of its broader mission—has developed a robust infrastructure to provide information about all people and businesses throughout the Nation.

#### ***Data-Driven Decisions: Choosing a Career***

*About 3.4 million students were expected to graduate from high school in 2013 (National Center for Education Statistics 2013). In addition, nearly 900,000 associate's degrees and 1.7 million bachelor's degrees were expected to be conferred in 2013 (National Center for Education Statistics 2014). These graduates, as well as adults considering a new job or a career change, need to make informed decisions about what career to pursue.*

*For more than 50 years, the Bureau of Labor Statistics' Occupational Outlook Handbook (OOH) has been a nationally recognized source of career information. It describes a variety of occupations, including what workers do on the job, working conditions, the training and education needed, earnings, and projected job growth. The handbook is available at <http://www.bls.gov/ooh/home.htm>.*

*The OOH relies on Government data, including the Current Population Survey, Census Bureau population projections, BEA's National Income and Product Accounts and Input-Output tables, BLS's National Employment Matrix, and Occupational Employment Statistics (Bureau of Labor Statistics, 2014b).*

*This information helps our youth throughout the country. For example, a college website advising prospective students on how to get ready for college suggests that 11th graders use the OOH to explore careers and their earning potential (Spoon River College, 2014). A high school honors student trying to decide on a career path used the OOH to help him decide to major in mechanical engineering. After graduation, he obtained a job at a large tech company (Loftus, 2008).*

- The Economic Census provides comprehensive information about the Nation’s commercial activity. Taken at 5-year intervals, it provides comparable data across economic sectors, using consistent time periods, concepts, definitions, classifications and reporting units. Larger firms report information directly, while administrative records are used to provide basic information for very small firms, reducing the need to burden them with questionnaires. In 2012, over 4 million businesses were contacted for information. The Economic Census provides information on over 29 million establishments, large and small, and together with the Census of Governments and the Department of Agriculture’s Census of Agriculture covers more than 98 percent of all economic activity.
- Data-Driven Decisions: Locating Hospitals***

*In 2013, \$41 billion was spent in the U.S. on health care construction (U.S. Census Bureau, 2013b), and Government data can be used to help improve the decisions regarding the location of this construction. For example, a multi-organizational committee led by Michigan’s Department of Community Health was set up in 2004 to revise its procedures for approving community hospitals. To help in this effort, it used data from the 2000 Decennial Census to identify areas with inadequate access to existing hospitals (Messina, et al., 2006).*
- The Population Census, conducted every 10 years, provides an official count of the entire population. Extensive efforts are directed to evaluating the coverage of the Census.
  - The American Community Survey (ACS), conducted every year, uses statistical samples of the population to provide more current and detailed information about local communities. In 2012, the ACS covered about 93 percent of the U.S. population.<sup>12</sup>
  - Other programs, such as the Quarterly Census on Employment and Wages and the Census of Fatal Occupational Injuries, both programs of the Bureau of Labor Statistics, collect and disseminate data that comprehensively covers the Nation.

When the Government collects data on a sample of persons or businesses (in order to avoid higher-cost collections involving the entire population) rigorous statistical methods are used to choose the sample that is representative of the relevant population as a whole, and the reliability of the estimates is specified through measures of sampling variability (Office of Management and Budget, 2006).

Private data collection—while perhaps produced more quickly—is often narrowly tailored to the specific needs of a business or industry. For example a private business might find it cost-effective to target specific “market segments” in its data collection efforts, such as persons with landline telephone access, or households above a certain income level. These choices make sense for that business’s specific needs, but the resulting data may be of limited use in determining whether the evidence has broader applicability.

The Government is also uniquely positioned to use administrative data to ensure coverage of the entire population while minimizing burden to respondents. Some data is already collected on nearly all employers and their employees for other purposes such as tax filings and unemployment insurance records. Data sharing arrangements enable Federal statistical agencies to use this information to construct survey frames and samples and to avoid redundancies in data collection, thereby minimizing respondent burden. Such arrangements are strictly governed by law and policy to protect the confidentiality of respondents, as discussed in the subsection on confidentiality below.

<sup>12</sup> Complete history of ACS coverage rates is available at: [https://www.census.gov/acs/www/methodology/coverage\\_rates\\_data/index.php#note2](https://www.census.gov/acs/www/methodology/coverage_rates_data/index.php#note2).

## 2. Consistent

Federal statistical agencies collect and disseminate data at regular intervals, and they work to ensure the data is consistent across data collection programs, enabling meaningful comparison and analysis. Government data collection is often broadly focused with a long time horizon, rather than targeted at specific products or specific interests in a short time horizon based on market demand. Government data collection efforts also focus on consistent definitions and measurements that make it easier to compare data across data sets. As a result, Government data collection facilitates comparisons with fewer discontinuities.

The long-term focus of Government statistical agencies facilitates the comparability of the data over time, allowing for the assessment of changes and trends. Although methodological changes and other factors can limit the strict comparability of some Government data series, Government responsibility for data collection and dissemination better ensures that the information collection activities will be ongoing and thus that the data will be consistent and comparable.

Government data collection efforts also facilitate comparisons of data across data sets. The Office of Management and Budget sets Government-wide definitions and classifications standards for concepts such as “race,” “ethnicity,” “Metropolitan Statistical Areas,” “industry,” and “occupation.” Without generally accepted definitions and standards, these concepts could be defined and measured in many ways, making it difficult to compare data from different sources.

In addition, the private sector might not find it profitable to provide or retain certain data for which there is not current adequate demand. To the extent that some of the value of data lies in meeting future demands that are not fully anticipated, as is often the case for basic scientific research, systematic collection (and archiving) of the data by the Government helps ensure that this future value of data is preserved.<sup>13</sup> For example, data from the Department of Agriculture about U.S. horse exports to Mexico was relatively unimportant for many years, but after U.S. horse-slaughtering policies changed in 2007, the time series of the data became important in understanding the effects of those policies (C-Fare, 2013, p. 7).

## 3. Confidential

Government statistical agencies recognize that the data they collect can be sensitive, and it is a top priority to maintain the confidentiality of the information. Many respondents would not provide accurate information if they believed that it would be made public or used to their detriment. Confidentiality is also the law: for example, the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA)<sup>14</sup> ensures that individuals and organizations who supply information for statistical purposes

### ***Data-Driven Decisions: Re-aligning a School District***

*There were 13,629 public school districts in the United States in the 2009-2010 academic year (National Center for Education Statistics, 2013). The value of public educational construction put in place in 2013 was \$63 billion (U.S. Census Bureau, 2013b). Government data can help improve decisions about public school district re-alignments and school construction. For example, the Fontana Unified School District in California experienced rapid population growth that required new schools to be built. In addition, changes in the distribution of population throughout the district created a need to review feeder patterns and re-align boundaries for all school levels. A consulting firm helped the district create new boundaries using analytic tools that rely on a wide range of Government data (DecisionInsite, 2013a), including data from the Census Bureau, BEA, the BLS, and other Government agencies (DecisionInsite, 2013b).*

<sup>13</sup> For examples of privately held data that is no longer available for future use, and the problems this poses for ongoing research, see <http://blogs.smithsonianmag.com/science/2013/12/the-vast-majority-of-raw-data-from-old-scientific-studies-may-now-be-missing/>.

<sup>14</sup> 44 U.S.C. § 3501 note (2002). See, titles 13 and 22 for confidentiality provisions specific to the Census Bureau and the Bureau of Economic Analysis, respectively.

under a pledge of confidentiality will have their identities protected and that the information will be used for statistical purposes only and may not be used for any enforcement or other regulatory purposes. Violators are subject to fines and incarceration.<sup>15</sup>

Private sector data collectors provide no similar guarantees; for example, a company that provides an Internet-based “do-it-yourself” survey application states that the anonymity of survey respondents’ information depends on the survey creator. “It is up to each survey creator to decide if they want to collect responses anonymously, or to capture respondents’ personal information.”<sup>16</sup> Similarly, social networking platforms are free to change privacy policies governing the use of data.

The Bureau of Economic Analysis (BEA) collection of data on services trade and on financial and operational information about investments abroad is a useful example of the value of the Government’s confidentiality protections. Because private firms have an incentive to safeguard information about internal operations and business plans, publicly reported information about their investment abroad and cross-border trade in services may be very limited or imprecise. This affects the usefulness of that information. BEA collects robust trade and financial data from private firms, yet publishes statistics based on the data without revealing anything that might violate individual firms’ confidentiality. This increases the odds that firms fully and accurately report the information.

#### **Data-Driven Decisions: Setting Contract Prices**

*The Producer Price Index (PPI) program at the Bureau of Labor Statistics (BLS) publishes monthly data on prices received by domestic producers of goods and services, allowing for the measurement of price changes from the seller’s perspective. Businesses can use the data to adjust prices in long-term sales and purchase contracts, allowing the parties to cope with changes in prices without otherwise revising their long-term business arrangements. These escalation clauses allow the parties to allocate risks in a mutually beneficial way over longer time periods, thereby reducing contract negotiation costs.*

*The data can have a large financial impact. According to a survey of Producer Price Index data users, the average contract that was escalated was valued at \$115 million (Bureau of Labor Statistics, 2013).*

#### **Data-Driven Decisions: Crime Mapping**

*Over 10 million crimes were committed in the United States in 2012, with property crimes representing the vast majority, about 9 million (Federal Bureau of Investigation, 2013). Crime mapping has emerged as a critical tool for law enforcement, pointing to relationships between various types of crime incidents and the characteristics of the neighborhood where the incidents occurred, such as population density, poverty, and housing vacancy. This helps to identify “hot spots” and focus law enforcement efforts where they are most needed. Crime mapping often relies on Census Bureau’s demographic and housing data, which is available at the neighborhood level (Reamer, 2006).*

*For example, the Illinois State Police department has used Census Bureau data to develop a list of risk factors to help identify methamphetamine hot spots. Of the eight risk factors that were ultimately developed, six of them came directly from Census Bureau data (Police Foundation, 2000).*

## **4. Credible**

Trust in the quality of Government statistics by data users is paramount if those statistics are to be valued. Statistical agencies foster credibility and trust in the veracity of their products by maintaining transparency and impartiality. When data is not transparent—for example if sources and methodology are unclear or undisclosed—questions arise about the quality and accuracy of the data.

<sup>15</sup> Section 513 of CIPSEA states that disclosure of confidential information is a class E felony punishable by imprisonment up to five years and fines up to \$250,000, or both.

<sup>16</sup> Additional explanation of the private survey company’s privacy and security policy is available at [http://help.surveymonkey.com/articles/en\\_US/kb/Are-my-survey-responses-anonymous-and-secure?](http://help.surveymonkey.com/articles/en_US/kb/Are-my-survey-responses-anonymous-and-secure?)

Government agencies strive to be transparent by providing full documentation of the survey forms used to collect the data and by publishing the statistical methodologies used to compile the data. Changes to statistical questionnaires require public notice and comment, which allows the public to provide the Government with important information that should be considered in the collection and dissemination of data. Agencies hold data user conferences and host online training in the use of their data in order to foster a better understanding of the strengths and limits of their data. By maintaining this high level of transparency, statistical agencies help the public understand where the statistics come from and what they represent, increasing the level of trust in the reliability of the data.

Credibility is also fostered by impartiality. The Office of Management and Budget (OMB) coordinates the Federal statistical system (44 U.S.C. 3504(e)) and issues policy directives that, among other things, aim to insulate agencies from pressures to manipulate data for political purposes or other special interest objectives. For example, OMB's Statistical Policy Directives Number 3 and 4 distinguish between policy-neutral data releases and interpretation of data by policy officials.<sup>17</sup>

## 5. Relevant

To produce relevant data, statistical agencies must balance the need to provide comprehensive data consistently over time against the need to be flexible in response to changes in the economy, society, and environment, as well as to changes in information needs. The rise of new technologies, industries, occupations, and household arrangements, shifts in policy priorities, and other societal changes may require changes in the kinds of data statistical agencies collect, the ways they collect them, the products they produce with them, and the vehicles they use to disseminate those products. Because of the strictures involved in changing Federal statistical agencies' data collection methods, private sector organizations, including academic and non-profit researchers, arguably have greater flexibility in collecting and classifying new kinds of data. To ensure they are collecting relevant data, statistical agencies frequently consult with private sector entities in the research and development of new data products.

### ***Data-Driven Decisions: Serving the Church Community***

*Houses of worship of various religions and denominations use information systems based on Census Bureau data to plan expansion initiatives, identify neighborhoods where a new location might be successful, and provide services to local communities based on who lives there. For example, a house of worship observing an increase in young families and a rising birthrate in the local area might decide to upgrade its nursery facilities and develop plans for summer camp and daycare activities (Leavell Center for Evangelism and Church Health, 2014). Often, houses of worship access this information through a third-party intermediary; for example, MissionInsite is an information and query tool that provides demographic information specifically tailored to churches (MissionInsite, 2014).*

Statistical agencies respond to changes in data needs by collecting new data or re-classifying existing data only after careful research and consideration. For example, U.S. statistical agencies work with agencies from other countries to develop new data categories for the National Income and Product Accounts and to conform to international definitions, methodologies, and standards (United Nations Statistics Division, 2014). Proposed changes in data collection and methodologies are carefully researched and reviewed by statistical agency staff, are vetted with statistical agency advisory committees and other statistical agencies that use the data, are posted in the Federal Register for comment by the public, and are approved by the Office of Management and Budget.

<sup>17</sup> See, 50 Fed. Reg. 38,932 – 38,934 (Sept. 25, 1985); and 73 Fed. Reg. 12,622 – 12,626 (Mar. 7, 2008).

Collection of data about household arrangements is a prime example of how statistical agencies alter data collection in response to societal changes. The classification of such household relationship data has evolved since 1880, when the Decennial Census questionnaire asked about the relationship of each person in the household to the householder (Simmons and O’Connell, 2003). In 1990, the category “unmarried partner” was added to the census to reflect the growing diversity of household arrangements and an increasing tendency for couples to live together before, or instead of, marrying. This category was added to other surveys over the next several years. Given continued changes in household and family relationships, there may be continued need to update or revise categories used in these surveys. OMB established an Interagency Work Group on Measuring Relationships in Federal Household Surveys to research the measurement of marital and household relationships and to make recommendations to statistical agencies for the development of questions that more accurately capture data on marriage and family arrangements (Office of Management and Budget, 2013, p. 63).

Another way statistical agencies can improve the relevance of their data is by re-organizing or re-classifying their data in ways that shed light on different sectors of the economy that were previously not clear. The Bureau of Economic Analysis (BEA) has developed “satellite accounts” in conjunction with its National Income and Product Accounts (NIPA) program as a means of responding to changing information needs. To provide better information about particular sectors of the economy, such as travel and tourism, BEA rearranges data elements underlying the official NIPA (Bureau of Economic Analysis, 2014a). In addition, BEA has also developed satellite accounts based on concepts that differ from the official NIPA concepts and serve as a framework for estimating vital sectors of the economy that were otherwise unmeasured (United Nations Statistics Division, 2009, p. 523-543). For example, until recently, the NIPAs treated spending on research and development as an expense, but BEA developed a satellite account that treated research and development as a capital investment, which is more consistent with the economic concept of research and development; in 2013 the research and development satellite account became a part of the official NIPAs (Bureau of Economic Analysis, 2013). BEA has also been developing a satellite account for health-care spending, which will reclassify health care spending as expenditures on treatment of various disease groups. The goal is to develop statistics that will provide perspective on overall health care spending as well as presenting a more detailed view of household consumption of care for diseases over time, thus helping researchers, businesses, and policy makers better assess the returns to society on medical-care spending (Bureau of Economic Analysis, 2014b).

### Box 5: Measuring E-Commerce: Government Data Sets the Standard After Building on Initial Private Sector Efforts

Until the late 1990s, only private sector organizations published estimates of the volume of sales transactions completed over a computer network, known as e-commerce. However, these private estimates ranged widely. For example, in 1998, private sector estimates of business-to-business e-commerce sales ranged from \$12 billion to \$92 billion, and estimates of business-to-consumer estimates ranged from \$4 billion to \$27 billion (Fraumeni 2001, p. 319).

The large differences in estimates likely reflected lack of uniformity in methodologies, coverage, and definitions, but such background information about the estimates was difficult to obtain or, if available, consisted of opaque descriptions on “sizing the Internet market” (Fraumeni 2001, p. 318).

During the summer of 1999, the Census Bureau developed detailed concepts and definitions of e-commerce in collaboration with other Government agencies and private sector experts (Parker and Grove 2000). The Bureau distinguished among three main components of e-business:

- E-business processes such as buying, selling, inventory and production management, logistics, and communication and support services.
- E-business infrastructure, including computers and other hardware; telecommunications and network channels; and system and applications software.
- E-commerce, which consists of transactions completed over a computerized network involving the transfer of ownership or right to use of goods and services—where the key determinant of when a transaction is “completed” is electronic agreement over the network and not payment.

The Census Bureau added e-commerce sales questions to its 1998 annual survey covering retail trade. The (revised) estimate for e-commerce retail sales in 1998 is about \$5 billion (U.S. Census Bureau, 2013a)—a little above the lowest private sector estimate for that year.

## 6. Accessible

Public access to Government data maximizes the potential beneficial uses of the data. In particular, providing access to data free of charge over the Internet helps ensure that data is used for more applications with positive benefits (net of other costs), particularly when the benefits are uncertain or diffuse. Equity in access to Government data ensures that all users have an equal opportunity to use the information.

Government data is becoming increasingly “open.”<sup>18</sup> Government data tabulations used to be disseminated exclusively in printed documents and made available for purchase or in libraries by the Government Printing Office; since the 1990s most Government statistical reports can be downloaded at no charge, while the data tabulations contained in those reports (and additional data) can be downloaded in electronic format. More recently, Government data has increasingly become available via Application Programming Interfaces (APIs) that allow programmers to access Government data directly and feed the

<sup>18</sup> As defined by the McKinsey Global Institute, fully open data is accessible to everyone; are machine readable (in formats that make it easy for computers to retrieve and process data); are offered at zero cost; and have no limits on reuse and redistribution (McKinsey Global Institute 2013, p. 3).



data into applications. President Obama signed an Executive Order on May 9, 2013, which promotes open data as the default mode for Federal Government information, and Office of Management and Budget issued a memorandum to establish a framework for institutionalizing the requirements of the Executive Order. Making machine-readable data available through APIs is one way statistical agencies are making large amounts of data available on a timely basis. Other efforts are underway to increase the amount of Government data that is open, including efforts at the National Oceanic and Atmospheric Administration, the National Aeronautics and Space Administration, the Department of the Interior, the Internal Revenue Service, and other agencies (U.S. Open Data Action Plan, 2014).

The Government does not necessarily have to make data more accessible on its own; the Government also supports private sector entities through grants, contracts, and public-private partnerships to make data more accessible and easier to use for the research community and the public at large. For example, the Minnesota Population Center of the University of Minnesota, under a grant from the National Institute of Child Health and Human Development, collects and preserves Current Population Survey microdata and documentation, harmonizes the data for compatibility over time, and disseminates the data at no charge (Minnesota Population Center, 2014). The Patent and Trademark Office, in conjunction with Google, made patent awards and applications more accessible and easier to search (McKinsey Global Institute, 2013, p. 12).

## The Complementary Relationship between Government and Private Sector Data

Government data is a vital source of information, but not the only important source. Government data and private data are complementary sources of information for decision making. In fact, even as private data production, aggregation, and dissemination have proliferated, much private data crucially depends on public data. For example, many private entities that create statistics, economic indicators, and other data products rely on Government data as a reference and standard for adjusting, weighting or testing the validity of their products. These entities include private sector entities that provide macroeconomic indicators that complement or attempt to forecast Government economic indicators; pollsters; and market research and consulting firms that collect and disseminate public opinion and marketing data.

For example, the Billion Prices Project is a private (academic) initiative to collect price data daily from hundreds of online retailers around the world in order to track inflation and prices. The data are commercially available through a private company, PriceStats, which uses the Consumer Price Index (CPI) as a benchmark to gauge the performance of the index. However, the PriceStats data is not intended to replace Federal Government data on prices as measured by the CPI, but instead provides an independent and complementary gauge of price changes (Lowrey, 2010).

In other instances, the role of the Government is to develop a “simple, reliable, and publicly accessible infrastructure that ‘exposes’ the underlying data” (Robinson, et al., 2009). The Government does so in a manner that is comprehensive, consistent, confidential, credible, relevant, and accessible, while the private sector is in the best position to develop custom-tailored analyses of Government data for specific users and specific purposes. That is, Government data can be used by the private sector as an input into more specialized products that the private sector is most suited to supply. Private sector entities, which are in a more direct relationship with customers, may be in a better position to perceive new or changing demands in the market. Private sector entities are freer to combine Government data from multiple Federal, State, local, overseas, and multinational agencies, as well as the private sector, and may be less restricted in presenting the data. The private sector has the flexibility to charge fees or sell advertising for data dissemination and therefore has a strong incentive to develop innovative ways to deliver Government data to the public. Moreover, the private sector is free to analyze and interpret Government data in any way it sees fit, subject to the discipline of the marketplace and peer-review. The flexibilities enjoyed by the private sector result in the ability to bring valuable products and services to the market.



## IV. The Commercial Value of Government Data

A primary goal of this report is to provide an initial framework for understanding the value of Government data. Government data is an input into the products and services of many types of businesses, so an important component of the value of Government data lies in its use as an input into these private sector products and services. Because Government data is disseminated freely and with no requirements for attribution, it is difficult to identify these uses of Government data and to estimate their monetary value. Unlike goods and services that are produced and sold in the marketplace, there is no market price for Government data, so it is not possible to use its price to reflect its value as a commercial input. Instead, this section examines the revenues attributed to firms that use Government data to illustrate the importance of Government data in the economy.

As a first step toward assessing the commercial value of Government data, we focus on defining the ecosystem of firms that rely heavily on Government data as a key input to their production processes. This section: (a) develops a definition of Government data-intensive business activities; (b) identifies some of the firms that engage in Government data-intensive business activities and estimates their size in terms of annual revenues; and (c) outlines a methodology for defining the Government data-intensive sector in terms of those industries<sup>19</sup> that most closely resemble Government data-intensive activities. The goal is to develop credible, albeit rough, estimates of the size of the sector.

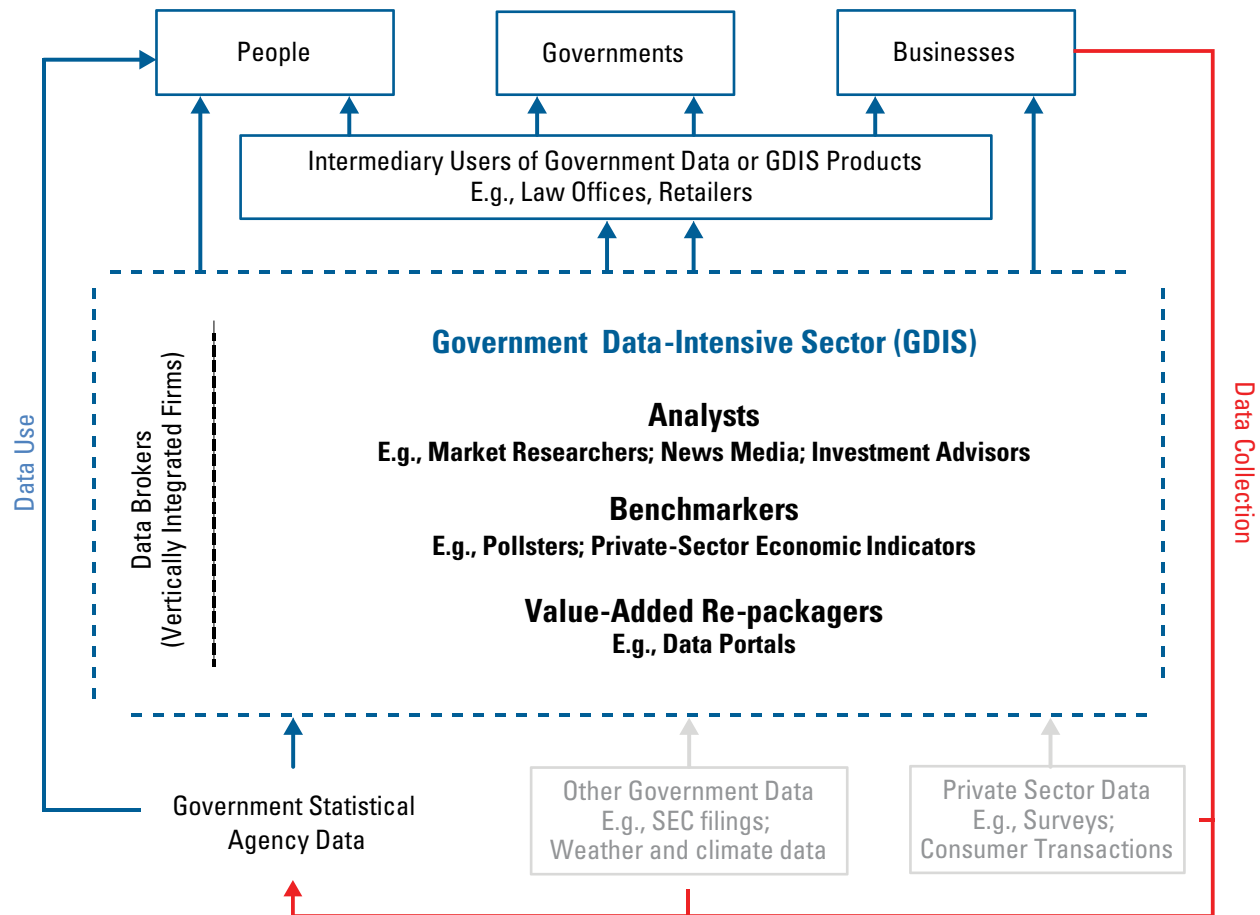
### Defining Government Data-Intensive Business Activities

This section provides a first step toward assessing the value of Government data as a commercial input by looking at the ecosystem of firms that rely heavily on Government data in their production processes. These firms constitute the Government data-intensive sector (GDIS). Government data-intensive businesses provide products and services to households, businesses, other organizations, and governments.

Figure 3 lays out where GDIS activities fall in a data value chain, with Government data flowing from Government data agencies (at the bottom of the diagram) to the ultimate users of the data (at the top of the diagram). While data can flow directly from the Government agencies to these users, data can also flow through the GDIS, which uses the data as an input for products and services purchased by the ultimate end users. (Note that in Figure 3 there are data flow arrows pointing in both directions, as Government both provides public data to users and collects raw data from them.) Below, GDIS activities are described in order from those furthest upstream in the data value chain to those furthest downstream.

<sup>19</sup> Ultimately the goal is to define the GDIS as a list of industries within the North American Industry Classification System (NAICS). See Appendix II for definitions of NAICS industry codes for Government data-intensive industries.

**Figure 3: Government Data Value Chain**



- **Value-Added Re-packagers** aggregate Government data from many different Federal Government agencies, private sector firms, other governments, and international organizations within a single application. The application adds value to the data by providing easy-to-use data interfaces and possibly the capability to create charts, tables, and maps, and by allowing users to switch periodicity easily (for example, switch between monthly, quarterly, and annual series). For-profit examples of re-packagers include Geolytics, Haver Analytics, and Socrata; non-profit examples include the Federal Reserve Economic Data (FRED) at the Federal Reserve Bank of St. Louis.
- **Benchmarkers** are firms that create statistics or other data products using non-Government data, but use Government data as a reference and standard with which to adjust, weight, or test the validity of their products. These firms include private providers of macroeconomic indicators that complement or attempt to forecast or substitute for Government economic indicators. Certain pollsters and market research firms that collect public opinion and marketing data and disseminate such information to private companies, marketing consulting firms, and news organizations also benchmark their survey results using Government data. Examples include Nielsen and Gallup.
- **Analysts** are firms that use Government data (either directly or indirectly through re-packagers) to create products for other firms or to generate research based on the data. For example, analysts create market segmentation products, in which households are classified into numerous categories

sharing similar tastes, incomes, household configurations, and interests. Market segmentation is a tool for more accurately and efficiently targeting resources devoted to media advertising buys, pricing strategies, and establishment locations. In addition, social science research organizations perform work under contracts from governments, foundations, and businesses and are often very intensive users of Government data. Yet another example is news organizations that report on the principal Government economic indicators and analyze their impact on the economy or that analyze other Government data, including highly localized demographic and economic information. Examples of news organizations that report Government data include CNBC, The New York Times Company, and the News Corporation.

- **Data Brokers** are firms that compile Government data (and sometimes proprietary private sector data) from around the world and add value by aggregating and integrating the data (McKinsey Global Institute, 2013, p. 6). These firms sell access to the resulting repackaged data and often provide analysis of the data and even news reporting on the data. Examples include Acxiom, Bloomberg, Esri, Experian, IHS Global Insight, and Thomson Reuters.

## Estimating the Size of the Government Data-Intensive Sector

As previously noted, estimating the value of Government data to GDIS firms is challenging. An ideal estimate of the commercial value of Government data to GDIS firms might be an estimate of how much lower the market value of GDIS firms would be absent Government data. However, it is not possible to get a counterfactual estimate of what the value of GDIS firms would be if there were no Government data available to them. Furthermore, many of the large GDIS firms are privately held, so there is no available information about their current market valuation.

Given these constraints, this report measures the size of the GDIS sector as a proxy for the commercial value of Government data. This subsection outlines a methodology for developing credible, albeit rough, estimates of the size and scope of the GDIS and reports such estimates. First, we identify specific companies that use Government data as an input into the products and services they sell. We describe the business activities of these companies and classify each company in one or more industries based on their business activities. The lower-bound estimate of the size of the GDIS is based on the revenues of the identified firms. The upper-bound estimate of the size of the GDIS is based on the revenues attributable to the industries identified by the assigned NAICS codes.

To begin, a partial list of companies known to use Government data to some degree was compiled based on consultations with statistical agency staff and lists of competitors cited in the 10-K filings or annual reports of publicly-traded companies. The firms highlighted in this section are selected for illustrative purposes, to provide a general sense of the business of firms operating in the GDIS, and no endorsement of their products and services is intended.

### **Data-Driven Decisions: Stabilizing Rents**

*There are over one million rent-stabilized apartments in New York City, regulated under the Local Emergency Housing Rent Control Act of 1962, the Local Rent Stabilization Law of 1969, and the Emergency Tenant Protection Act of 1974. These State and city laws require that data be collected every three years to determine whether rent controls should continue. The purpose of the New York City Housing and Vacancy Survey (NYCHVS) is to provide such data. The data includes the rental vacancy rate, the supply of housing accommodations, and the condition of such accommodations. Since 1965, the City has retained the Census Bureau to design and carry out the NYCHVS in order to fulfill the data collection requirement. The latest NYCHVS survey in 2014 is estimated to cost approximately \$7.8 million, which is a small investment relative to the billions of dollars of foregone market rents at stake.*

Below we briefly describe the business activities of these firms, along with the industry or industries in which each firm operates. We classify each firm in one or more industries based on the description of the business activities. Appendix II lists and describes the industries (as defined by their 6-digit North American Industrial Classification System (NAICS) codes).

**Acxiom**, headquartered in Little Rock, Arkansas, is a large data broker. One of its services is a market segmentation tool called PersoniX. Clients use PersoniX to refine products to meet the needs and desires of particular groups of potential customers; to formulate marketing strategies; and to improve targeting of advertising (Acxiom, 2014). PersoniX classifies households into 70 groups with similar “needs, motivations, likes and dislikes and behaviors” within 21 “life stage groups” (households that have attained similar life milestones, such as getting married, having a child, or buying a home).

PersoniX draws on its analysis of Acxiom’s database of demographic, contact, and purchasing behavior information for 125 million U.S. households. It gathers data from “various public information sources: driver’s license files, birth records, self-reported surveys, county tax assessors, telephone books and other public sources of data” such as Census Bureau data (Acxiom, 2010; Ayres, 2007, pp. 145-6).

*NAICS Industries: Marketing Consulting Services (NAICS 541613)*

### ***Data-Driven Decisions: Locating a Business Establishment***

*According to the Census Bureau’s Statistics of U.S. Businesses, about 650,000 new establishments opened in 2009-2010, employing 4.9 million workers. Businesses use Government data to make informed decisions about where best to locate new establishments. For example, PETCO, a large pet supplies retailer, decided to implement a geographic information systems (GIS) site selection tool. The goal was to mitigate the risk associated with opening new locations and to maximize the return on invested capital in new stores. The tool, ESRI Business Analyst, which uses demographic data from the 2010 Decennial Census and the Census Bureau’s American Community Survey, also helped the company analyze geographic decisions related to merchandise planning and advertising. The company found that using the site selection tool reduced its real estate risk (Pillar, 2010).*

**Bloomberg, L.P.** is a data broker that is the world’s biggest provider of financial information. The company’s main product (accounting for 85 percent of the company’s revenue) is its data terminal, which the company leases to its customers. The terminal provides 15,000 functions over the Internet, including Government data series as well as securities market data and other news plus analytical and computational capabilities (Elkind 2013). Another product published on the Internet, Bloomberg Government, provides many Federal Government documents and data (for example, the Federal Register, reports published by the Government Accountability Office, and data on Federal Government contracts and grants) and adds value by providing sophisticated search tools. Bloomberg Government also provides detailed profiles of Congressional districts using Census Bureau and other data.

*NAICS Industries: News Syndicates (NAICS 519110); Internet Publishing and Broadcasting and Web Search (NAICS 519130)*

**Environmental Systems Research Institute (Esri)** provides software tools and data that enable customers to produce maps by combining geographic information and also to help solve problems that have a spatial element to them. (For example, we describe in a sidebar elsewhere in this report an example of a business using Esri’s Business Analyst tool to decide where to locate.) The base maps that Esri makes available to map-makers using Esri’s flagship software product, ArcGIS, were developed using geographic data from numerous sources, including Federal Government agencies. For example, Esri’s World Topographic Map draws on geographic data from the U.S. Geological Survey, the Environmental Protection Agency, and the U.S. National Park Service (Environmental Systems Research Institute, 2014a). Esri also packages, adds value, and sells economic and demographic data by geographic location from the 2010 Decennial Census and the American Community Survey that can be combined with base maps to generate maps that reflect the geographic distribution of the data (Environmental Systems Research Institute, 2014b). The same Census Bureau data is used by Esri in the development of its Tapestry market segmentation tool (Environmental Systems Research Institute, 2014c); Tapestry categories are matched by Esri to data from the Consumer Expenditure Survey conducted by the Bureau of Labor Statistics to analyze consumption patterns by product class and market segment (Environmental Systems Research Institute, 2014d).

*NAICS Industries: Internet Publishing and Broadcasting and Web Search Portals (519130); Marketing Consulting Services; Marketing Research and Public Opinion Polling (541910); and Administrative Management and General Management Consulting Services (541611)*

**Gallup** collects data on the economic status, attitudes, wellbeing, and behaviors of individuals in more than 160 countries (Gallup, 2013a; Gallup, 2013b). In the United States, Gallup uses random-digit-dialing of landline and mobile phone numbers to generate samples of individuals to survey. After it collects and processes its survey data, Gallup assigns weights to each respondent so that the demographic profile of the weighted sample matches the demographic profile of the entire adult population. The weights are based on Census Bureau data from the American Community Survey, the Current Population Survey, and the Decennial Census on gender, race, age, educational attainment, and region. (Gallup 2010; Gallup, 2013c). Gallup publishes many of its findings on its website, and many news organizations also report its findings. Gallup Analytics makes detailed poll findings available to subscribers for an annual fee; for an additional fee subscribers can gain access to the respondent data (Gallup, 2013b).

*NAICS Industries: Marketing Research and Public Opinion Polling (NAICS 541910)*

**GeoLytics, Inc.** repackages Decennial Census, American Community Survey data from the Census Bureau and Uniform Crime Reporting Program Data from the U.S. Dept. of Justice, Federal Bureau of Investigation and adds value by providing software to make the data easier to access, analyze, and map (GeoLytics 2014a and 2014b). It produces products such as Business Demographics 2009 “for making sound business decisions” that combines Decennial Census population data with data from the Economic Census and County and Zip Code Business Patterns data as well as data from the Consumer Expenditure Survey of the Bureau of Labor Statistics and includes analytic and mapping software and GeoLytics population projections (GeoLytics 2014c). GeoLytics also sells market research data, including a market segmentation product, based on Census Bureau demographic data (GeoLytics 2014d).

*NAICS Industries: Internet Publishing and Broadcasting and Web Search Portals (519130); Marketing Consulting Services; Marketing Research and Public Opinion Polling (541910); and Administrative Management and General Management Consulting Services (541611)*

**GreatSchools** is a national non-profit organization with offices in San Francisco, Milwaukee, Washington D.C., and Indianapolis. GreatSchools ratings are based on data drawn from State Departments of Education on the most recent available standardized test results for schools in each State. For each grade/subject combination, GreatSchools determines ratings based on which decile the school's test scores are in the statewide distribution of test results. For each grade the subject-level ratings are averaged to arrive at an overall rating for that grade. Similarly, school-wide and district-wide ratings are weighted (by the appropriate school population) averages of grade-level and school-level ratings, respectively.

*NAICS Industries: Internet Publishing and Broadcasting and Web Search (NAICS 519130)*

**Haver Analytics** updates and manages historical time series data, providing up-to-the-minute data on over 200 economic and financial databases to customers in the global strategy and research community. The company maintains offices in New York, London, and Singapore and serves clients in financial services, government, academia and various industry groups from consulting to manufacturing. Data offerings range from daily market data to annual economic statistics, with extensive forecast data covering the world's economies. Source data includes over 1,200 government and private sources. Haver databases are available through specialized software that allows for database integration and analysis, and Haver also offers web delivery of data. Telephone client support is staffed by database managers who can help clients locate and understand the data available to meet their information needs (Haver, 2014).

*NAICS Industries: Administrative Management and General Management Consulting Services (NAICS 541611)*

### **Data-Driven Decisions: Investing in Real Estate**

*Existing-home sales in 2013 were 5.1 million on a seasonally adjusted annual rate according to the National Association of Realtors, and real estate investors accounted for roughly 20 percent of home purchases near the end of 2013, according to the Campbell/Inside Mortgage Finance survey of real estate conditions (Bharatwaj, 2013). Both investors and individual home buyers can use Census Bureau data to help make better residential real estate investment decisions. For example, a real estate investor used Government data from the Census Bureau's "State & County QuickFacts" website (<http://quickfacts.census.gov/qfd/index.html>) to help to determine whether houses were under- or overpriced in a given county. By his rule of thumb, houses were underpriced in that county, so he bought houses that he later sold at a higher price. He recommends that prospective real estate investors consider Census Bureau data such as household income, population growth, unemployment rate, job growth, housing inventory, rental rates, and median home price in making investment decisions (Pierce, 2009).*



**IHS Inc.** is a publicly traded business information services company based in Englewood, Colorado, with experts in energy, economics, geopolitical risk, sustainability and supply chain management. The company was originally founded in 1959 as a provider of product catalog databases on microfilm for aerospace engineers, and it now employs over 8,000 people in more than 31 countries. IHS provides business clients with global market and economic information, including technical information, tools and operational and advisory services. IHS Data & Analytics services provide guidance in the areas of economics and finance, risk managements, country risk and credit risk, investment and portfolio analysis and management, corporate strategy, marketing and market intelligence, business development, and insurance underwriting. Customers range from governments and large multinational corporations to small companies and technical professionals (Ycharts, 2014).

*NAICS Industries: Marketing Consulting Services (NAICS 541613); Administrative Management and General Management Consulting Services (541611)*

---

**MapQuest, Inc.** offers online, mobile, business and developer solutions that help people discover and explore where they would like to go, how to get there and what to do along the way and at the destination. MapQuest's suite of mobile services extends the company's popular place search, mapping and directions services beyond the desktop to cell phones and other wireless devices. MapQuest Mobile offers free solutions for the mobile web as well as free voice-guided navigation applications to meet the needs of MapQuest users on-the-go. Recently, to provide "a more robust local discovery experience," MapQuest has included additional information from a variety of sources, including publicly available Government data. For example, it includes airport status information by the Federal Aviation Administration; Mapquest reports that airports are the top queried search (AOL, 2014)

*NAICS Industries: Internet Publishing and Broadcasting and Web Search (NAICS 519130)*

---

**Nielsen** is a global marketing and advertising research company headquartered in Lower Manhattan, New York City. It offers critical media and marketing information, analytics, and industry expertise about what consumers buy and what consumers watch. Customers include media, entertainment, and consumer goods companies. Nielsen is active in over 100 countries, and employs approximately 32,000 people worldwide. Total revenues amounted to \$5.1 billion in 2010 (Ycharts, 2014).

The company relies on Government data to adjust the results of its information collection. For example, Nielsen calculates the number of homes with televisions and the number of television viewers in the United States. To do so, Nielsen uses proprietary data from its People Meter panel, along with Census Bureau data and auxiliary sources such as state governments and the U.S. Postal Service. Nielsen applies TV penetration rates to convert the total household and population estimates to TV households and persons therein (Nielsen, 2013).

*NAICS Industries: Marketing Research and Public Opinion Polling (NAICS 541910)*

**Thomson Reuters** provides “intelligent information” for businesses and professionals by combining industry expertise to deliver critical information to decision makers in the financial and risk, legal, tax and accounting, intellectual property and science and media markets. The company’s Financial & Risk Division provides solutions to the global financial community, delivering news, information and analytics to facilitate transactions relationship among trading, investing, financial and corporate professionals.

*NAICS Industries: News Syndicates (NAICS 519110), Administrative Management and General Management Consulting Services (NAICS 541611)*

---

**Trulia** is an on-line search engine with a database of 3.5 million homes (Trulia, 2014a). It delivers information to homebuyers, sellers, owners, and renters about properties (including mapping homes for sale, rent, and recently sold for a chosen locality), agents and neighborhoods around the country (Trulia, 2014b). Neighborhood information includes data on schools (from GreatSchools—see description above). Crime hotspots are mapped using data provided by SpotCrime.com and CrimeReports.com. Trulia reports community statistics based on 2000 Decennial Census data for a locality with comparison data for its county and State. These statistics include median household income and the distributions of home ages, household income, and commuting times. The Trulia website also provides numerous links to other sources of data on a particular community, including Bureau of Economic Analysis data on county-level per capita personal income and its components and Census Zip Code Business Patterns by NAICS industry code (Trulia, 2014c).

*NAICS Industries: Other Activities Related to Real Estate (NAICS 531390)*

---

**Zillow** is an online real estate listing service headquartered in Seattle with offices in New York City, San Francisco, Chicago, Irvine, CA and Lincoln, NE.

Zillow’s website draws on a continually updated database of 110 million U.S. homes currently for sale or rent or not on the market and maps them for a chosen locality. The data on listed homes is drawn from public data on physical attributes, tax assessments, and prior and current transactions, which can also be updated by home sellers (Zillow, 2014a). Zillow also offers “Zestimates” for home values and rents using a proprietary automated valuation model based on Zillow’s database (Zillow, 2014b). Zillow also provides local and (for comparison) national demographics data from the 2000 Decennial Census, including data on relationship status, homes with children, age distribution, median household income, percent population that consists of single males and single females, median age, average household size, and average commute time compared with national numbers. In addition, using segmentation methods based on 2000 Decennial Census data, Zillow reports on a locality’s “main types of people.” Zillow also maps school catchment areas and reports ratings of local public schools using proprietary data from GreatSchools (Zillow, 2014d); see description of GreatSchools above.

*NAICS Industries: Other Activities Related to Real Estate (NAICS 531390)*

## 1. Lower-Bound Estimate of Size of the GDIS Based on Size of Selected Firms Active in the GDIS

Table 3 lists the companies identified as relatively intensive users of Government data, along with their revenues. We use the revenues of these firms as the basis of the lower bound estimate of GDIS revenues. The sources of revenues estimates for these firms were Gale Business Insights and OneSource by InfoGroup.<sup>20</sup> Although this list is far from complete, it is useful to the extent that it identifies revenues attributable to some businesses that create and sell products using Government data, and thus it represents a lower-bound estimate of the value of Government data as a commercial input.

It is important to note that, even if Table 3 were a complete list of GDIS firms, their total revenues can only be a rough estimate of the value of Government data. The wide variation in the intensity with which companies use Government data to produce other goods and services presents another difficult challenge to estimating the value of the Government data they use as an input. For a handful of value-added re-packagers, the bulk of their activities may be so closely associated with the use of Government data that it is almost incontrovertible that most of their revenues are a reflection of the value of Government data. For other firms, the picture is less clear—Government data is an input into their production processes (either directly or indirectly via re-packagers and brokers), but the extent to which that is the case is unclear. Moreover, there are government-supported and non-profit entities, such as the National Bureau of Economic Research and the Minnesota Population Center, that provide similar services for free. Nonetheless, the revenues of firms that are relatively intensive users of Government data reflect, at least to some extent, the commercial value of the Government data they use. While future work may provide more refined estimates, these initial estimates at least illustrate the important point that the Nation's investment in Government data is small relative to the revenues of the firms that rely on the data to produce goods and services for economic investors.

<sup>20</sup> We also considered the NAICS industry codes that these two databases reported for each company, but we did not find them to be accurate based on the companies' own descriptions of their activities on their websites.

**Table 3: Revenues and Employment of Firms in the Government Data-Intensive Sector**

Company	Listed/Unlisted/Subsidiary	Revenues	Employment
Acxiom	L	\$1,130,624,000	6,300
Bloomberg	U	7,560,000,000	15,000
Esri	U	1,030,470,000	2,900
Gallup	U	121,185,000	2,700
Geolytics	U	4,400,000	10
GreatSchools	U	N/A	N/A
Haver Analytics	U	11,000,000	38
IHS Global	L	1,530,000,000	6,000
MapQuest	S	45,210,000	120
Nielsen	L	5,407,000,000	35,000
Thomson Reuters Financial & Risk Division	L (Canada)	7,193,000,000	20,700
Trulia	L	68,085,000	519
Zillow	L	117,000,000	560
<b>TOTAL</b>		<b>\$24,217,974,000</b>	<b>89,847</b>

*Notes:* Table revised after initial publication. Data are for 2012 except for Bloomberg (2011) and Geolytics (2010).

Listed companies are publicly traded companies; Unlisted companies are not publicly traded; and Subsidiary companies are subsidiaries of publicly traded companies. N/A is not available.

*Sources:*

- Gale Business Insights: ESRI, Bloomberg, Geolytics
- OneSource (Infogroup): Haver, Mapquest, Gallup, IHS, Trulia, Zillow.
- Annual Reports: Acxiom, Nielsen, Thompson Reuters.

## 2. Upper-Bound Estimate of the Size of the GDIS Based on the Size of Government Data-Intensive Industries

Using the industry classification of each company in Table 3, we identify industries to include in the GDIS sector. Table 4 lists those industries (by 6-digit NAICS code) and their size, based on Census Bureau estimates of the revenues of employer firms in these industries.

**Table 4: Revenues of Government Data-Intensive Sector (GDIS) by Industry**

NAICS Code	Industry Name	2012 Revenues (\$billions)
519110	News Syndicates	\$2.3
519130	Internet Publishing and Broadcasting and Web Search	57.5
531390	Other Activities Related to Real Estate	12.0
54161	Management Consulting Services <sup>21</sup>	131.7
541910	Marketing Research and Public Opinion Polling	17.3
<b>Total</b>		<b>\$220.8</b>

Source of revenues data: U.S. Census Bureau. 2013d. Revenues are for employer firms only.

It is important to note that there is no bright line that separates the NAICS codes that compose the GDIS from the rest of the economy.<sup>22</sup> An alternative approach to defining the GDIS has been taken by the Open Data 500 project, which is the first comprehensive study of U.S. companies that use open Government data to generate new business and develop new products and services and conducted by the GovLab at New York University. Open Data 500 has compiled a list of 500 U.S. firms and organizations that use Government data based on the Open Data 500's outreach, expert advice, and research (Open Data 500, 2014). In many cases, the 17 company categories used by Open Data 500 also do not neatly map to a well-defined list of NAICS codes.<sup>23</sup> Moreover, in both our GDIS classification system and that of Open Data 500, it is difficult to determine precisely the intensity with which certain firms or industries use Government data. Undoubtedly there are some firms in the GDIS that do not use Government data, while there are others that do use Government data at least to some degree but are classified in industries that are excluded

<sup>21</sup> NAICS code 54161 includes these industries (for which we don't have separate revenue estimates):

- 541611 Administrative Management and General Management Consulting Services
- 541612 Human Resources Consulting Services
- 541613 Marketing Consulting Services
- 541614 Process, Physical Distribution, and Logistics Consulting Services
- 541618 Other Management Consulting Services

<sup>22</sup> Note that there are no official NAICS codes for GDIS or any set of other data-driven industries for that matter. For example, a Government Accountability Office report on "information resellers," which are "companies with a primary line of business of collecting, aggregating, and selling personal information to third parties" notes that there is no well-defined NAICS code for such companies (U.S. Government Accountability Office, 2013 pp. 4-5).

<sup>23</sup> The categories of companies used by Open Data 500 are:

- Aerospace and Defense
- Business & Legal Services
- Data/Technology
- Education
- Energy
- Environment and Weather
- Finance & Investment
- Food & Agriculture
- Geospatial/Mapping
- Governance
- Healthcare
- Housing/Real Estate
- Insurance
- Lifestyle & Consumer
- Research & Consulting
- Scientific Research
- Transportation

from our definition of the GDIS. Other attempts to estimate the size of other data-intensive sectors, such as the economic impact of the “geo-services sector” (Oxera 2013) or the size of the “data-driven marketing economy” (Deighton and Johnson 2013), also encountered challenges in parsing these sectors from the rest of the economy.

Nevertheless, even given the caveats to the estimates in Tables 3 and 4, they provide a rough order of magnitude of the size of the Government data-intensive sector. The lower-bound estimate, based on a very short and incomplete list of firms that rely on Government data, suggests that Government data helps private sector firms generate revenues of at least \$24 billion annually. The upper-bound estimate suggests that GDIS industries generate revenues of \$221 billion. While there is clearly a large range of the estimate of the commercial value of Government data as measured in terms of the GDIS, the important conclusion is that Government data supports an important sector of the economy that itself supports myriad business investment decisions across the globe.

## V. Conclusion

Historically, data collection and dissemination have been the purview of Government. Information, including statistics about the Nation's economy and people, has been considered a public good. Therefore, providing such information was seen as an appropriate function of Government, helping to ensure that the optimal quantity and quality of data would be produced.

Focusing on Federal statistical data in particular, this report outlines evidence to suggest that Government data provides important information that could not or would not be produced otherwise. Given the market failures that are likely to occur in the private production of information, Government data production results in more information than would otherwise be available. Moreover, the high quality of Federal statistical data—as summarized by its comprehensiveness, consistency, confidentiality, credibility, relevance, and accessibility—results in better information than would otherwise be available. This information supports decision makers throughout the Nation, including individuals, businesses, other institutions, and governments at every level.

Government data and private data are complementary. The high quality of Government data enhances the value of private sector data by providing the benchmark that lends context and validity to private sector data. In addition, the private sector uses Government data to create new products and services in a wide range of industries. While its commercial value represents only one aspect of the value of Government data, it nonetheless provides a sense of the importance of Government data in the economy. This report estimates the commercial value of Government data using a rough estimate of the size of the Government data-intensive sector (GDIS)—the private industries that rely on Government data as a critical input to the goods and services they produce. It highlights specific examples of how firms in GDIS industries create value using Government data, including Federal statistics.

The value of the Nation's investment in Government data lies in the usefulness of the data in improving the countless decisions it supports, either directly or indirectly through the products and services of the private industries that rely on Government data. Estimating this value would require counter-factual evidence about the decisions that would be made without Government data. Such evidence is currently not available but represents a promising area of research.





## References

- Acxiom. 2010. *PersonicX Cluster Perspectives*. (reference.mapinfo.com/software/anysite\_segmentation/english/2\_0\_1/PersonicX\_Binder.pdf accessed January 2, 2014).
- Acxiom. 2014. "PersonicX." (<http://www.myacxiom.com/> accessed January 2, 2014).
- Acxiom. 2014. *Acxiom Corporation Annual Report 2013*. (<http://d3u9yejw7h244g.cloudfront.net/wp-content/uploads/2013/09/2013-Annual-Report.pdf> accessed June 3, 2014).
- Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *The Quarterly Journal of Economics*, Vol. 84, No. 3, pp. 488-500.
- Allen, Beth. 1990. "Information as an Economic Commodity." *American Economic Review*. 80:2(May), pp. 268-273.
- America Online. "MapQuest Overhauls Local Offering with Multiple Data-service Partnerships and Integrations," April 9, 2014 (<http://corp.aol.com/2014/04/09/mapquest-overhauls-local-offering-with-multiple-data-service-par> accessed on June 2, 2014).
- American Planning Association. 2014. (<https://www.planning.org/apaatagance/history.htm> accessed on May 8, 2014).
- Ayres, Ian. 2007. *Super Crunchers: Why Thinking-by-numbers is the New Way to be Smart*. Bantam. (<http://books.google.com/books?id=WzhgbJ59z18C&pg=PA146&dq=acxiom+%22census+data%22&hl=en&sa=X&ei=HNLFUqyUNrSusASK9YG4DQ&ved=0CEIQ6AEwAA#v=onepage&q=acxiom%20%22census%20data%22&f=false> accessed January 2, 2014).
- Bates, Benjamin J. 1988. "Information as an Economic Good: Sources of Individual and Social Value." In *The Political Economy of Information*. Ed. Vincent Mosco and Janet Wasco. (pp. 76-94). Madison WI: University of Wisconsin Press.
- Bharatwaj, Shanthi. 2013. "Investor Share of Home Purchases Continues to Tick Up." (December 23). (<http://housingpulse.com/> accessed May 8, 2014).
- Boisot, Max and Agustí Canals. 2004. "Data, Information, and Knowledge: Have We Got It Right?" *Journal of Evolutionary Economics*. 14: 43-67.
- Boston Consulting Group. 1991. "New BCG Study Re-evaluates Size, Growth and Importance of Business-to-Business E-Commerce." *PR Newswire* (December 21). (<http://www.thefreelibrary.com/New+BCG+Study+Re-evaluates+Size,+Growth+and+Importance+Of...-a058317733> accessed January 21, 2014).
- Bureau of Economic Analysis. 2013. "How Did BEA Change the Treatment of Spending For Research and Development During the 2013 Comprehensive Revision?" ([http://www.bea.gov/faq/index.cfm?faq\\_id=1028](http://www.bea.gov/faq/index.cfm?faq_id=1028) accessed January 7, 2014).
- Bureau of Economic Analysis. 2014. *National Income and Product Accounts*. Table 5.1. Saving and Investment by Sector. (<http://bea.gov/iTable/iTable.cfm?ReqID=9&step=1#> accessed February 21, 2014).
- Bureau of Economic Analysis. 2014a. "Information Guide: Travel and Tourism." (<http://www.bea.gov/industry/iedguide.htm#TTSA> accessed January 7, 2014).

- Bureau of Economic Analysis. 2014b. "Health Care Satellite Account." ([http://www.bea.gov/national/health\\_care\\_satellite\\_account.htm](http://www.bea.gov/national/health_care_satellite_account.htm) accessed January 7, 2014).
- Bureau of Economic Analysis. 2014c. "Security and Release Procedures." (<http://www.bea.gov/about/pdf/security-release-procedures.pdf>.)
- Bureau of Labor Statistics. 2013. "Highlights of the 2013 PPI User Survey," Beyond the Numbers. Volume 2, No. 20. (<http://www.bls.gov/opub/btn/volume-2/highlights-of-the-2013-ppi-user-survey.htm> accessed July 7, 2014).
- Bureau of Labor Statistics. 2014a. *National Longitudinal Survey of Youth 1979*. "Number of Jobs Held by Individuals from Age 18 to Age 46 in 1978 to 2010 by Educational Attainment, Sex, Race, and Hispanic or Latino Ethnicity." (<http://www.bls.gov/nls/nlsy79r24/jobsbyedu.pdf> accessed February 27, 2014).
- Bureau of Labor Statistics. 2014b. Employment Projections Methodology. ([http://www.bls.gov/emp/ep\\_projections\\_methods.htm](http://www.bls.gov/emp/ep_projections_methods.htm) accessed February 27, 2014).
- Center City District. 2014. "About Us" (<http://www.centercityphila.org/about/CCD.php> accessed February 27, 2014).
- "Council on Food, Agriculture, and Resource Economics (C-FARE): 2013 Annual Report." 2014. (<http://www.cfare.org/UserFiles/Reports/2013AnnualReportFINAL-2-26-2014.pdf> accessed on February 27, 2014).
- Cornes, Richard and Todd Sandler. 1996. *The Theory of Externalities, Public Goods, and Club Goods*. New York: Cambridge University Press.
- Deighton, John and Peter A. Johnson. 2013. *The Value of Data: Consequences for Insight, Innovation & Efficiency in the U.S. Economy*. (Funded by the Data-Driven Marketing Institute of the Direct Marketing Association. <http://ddminstitute.thedma.org/#valueofdata> accessed July 7, 2014).
- DeLong, J. Bradford and A. Michael Froomkin. 2000. "Speculative Microeconomics for Tomorrow's Economy." *The Economics of Digital Information and Intellectual Property*. Cambridge: MIT Press.
- Duncan, Joseph W. and William C. Shelton. 1978. *Revolution in the United States Government Statistics 1926-1976*. U.S. Dept. of Commerce Office of Federal statistical Policy and Standards.
- Elkind, Peter. 2013. "The Trouble at Bloomberg." *Fortune*. (December 23), pp. 130-147.
- Environmental Systems Research Institute (ESRI 2014a). May 2014. "ArcGIS: World Topographic Map." (<http://www.arcgis.com/home/item.html?id=30e5fe3149c34df1ba922e6f5bbf808f> accessed May 28, 2014).
- Environmental Systems Research Institute (ESRI 2014b). May 2014. "Demographic, Consumer, and Business Data: Demographics." ([http://www.esri.com/data/esri\\_data/demographic-overview](http://www.esri.com/data/esri_data/demographic-overview) accessed May 28, 2014).
- Environmental Systems Research Institute (ESRI 2014c). March 2013. *Methodology Statement: Esri Data—Tapestry™ Segmentation*. (<http://www.esri.com/~media/Files/Pdfs/library/whitepapers/pdfs/esri-data-tapestry-segmentation.pdf> accessed May 28, 2014).
- Environmental Systems Research Institute (ESRI 2013d). August 2013. *Methodology Statement: 2103 Esri US Consumer Spending Data*. (<http://www.esri.com/~media/Files/Pdfs/library/whitepapers/pdfs/consumer-expenditure-methodology-2013.pdf> accessed May 28, 2014).
- European Commission. 2011. *Pricing of Public Sector Information Study (POPSIS)*. (<http://www.epsiplatform.eu/content/popsis-assessment-psi-charging-policies> accessed July 7, 2014).

- Export.gov. 2014. "Success Stories: Coffee & More, LLC." ([http://export.gov/articles/successstories/eg\\_success\\_story\\_022775.asp](http://export.gov/articles/successstories/eg_success_story_022775.asp) accessed February 27, 2014).
- Federal Bureau of Investigation. 2013. "Table 1: Crime in the United States 2012." (<http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2012/crime-in-the-u.s.-2012/tables/1tabledatadecoverviewpdf/tab1overview.pdf> accessed July 7, 2014)
- Fraumeni, Barbara M. 2001. "E-Commerce: Measurement and Measurement Issues." *American Economic Review*. 91:2 (May), pp. 318-322.
- Gallup. 2010. "How Are Polls Conducted?" (<http://www.gallup.com/poll/File/125927/How%20Are%20Polls%20Conducted%20FINAL.pdf> accessed January 2, 2014).
- Gallup. 2013a. "About Gallup." (<http://www.gallup.com/corporate/115/about-gallup.aspx?ref=f> accessed January 2, 2014).
- Gallup. 2013b. "Gallup Analytics." (<http://www.gallup.com/gallupanalytics.aspx?ref=b> accessed January 2, 2014).
- Gallup. 2013c. Telephone Conversation with Stephanie M. Kafka, Consultant and Survey Methodologist (December 6).
- GeoLytics. 2014a. "Demographic Data." (<http://www.geolytics.com/resources/demographic-data.html> accessed May 28, 2014.)
- GeoLytics. 2014b. "Crime Reports." (<http://geolytics.com/USCensus,Crime%20Reports,Data,Features,Products.asp> accessed May 28, 2014).
- GeoLytics. 2014c. "Business Demographics 2009." (<http://www.geolytics.com/resources/demographic-data.html> accessed May 28, 2014).
- GeoLytics. 2014d. "Market Research Data." (<http://www.geolytics.com/resources/market-research-data.html> accessed May 28, 2014).
- Gilchrest, Lauren. 2013. "Defining Downtown Using LED." (<http://lehd.ces.census.gov/doc/workshop/2013/Defining%20Downtown%20-%20Lauren%20Gilchrist%206-13-13.pdf> accessed July 7, 2014).
- GreatSchools. 2014. "GreatSchools Ratings FAQ." (<http://www.greatschools.org/find-a-school/defining-your-ideal/2423-ratings.gs#7> accessed January 3, 2014).
- Haver Analytics. 2014. ([www.haver.com](http://www.haver.com) accessed May 30, 2014).
- Hirshleifer, Jack. 1972. "The Private and Social Value of Information and the Reward to Inventive Activity." *American Economic Review*. 61:561-574.
- Konkel, Frank. 2013. "NOAA to Seek Partnership to Boost Data Sharing." *Federal Computer Weekly*. (December 19) (<http://fcw.com/Articles/2013/12/19/NOAA-data-RFI.aspx> accessed January 5, 2014).
- Lazo, J.K., Morss, R.E., and J.L. Demuth. 2009. "300 Billion Served: Sources, Perceptions, Uses, and Values of Weather Forecasts." *Bulletin of the American Meteorological Society*. 90(6).
- Leavell Center for Evangelism and Church Health. 2014. "Demographic Profiles." (<http://www.leavellcenter.com/Demographic%20Services/Demographic%20Profiles.asp> accessed January 22, 2014).

- Loftus, Margaret. 2008. "Untapped Potential" *ASEE Prism*, Vol. 18, No. 2. ([http://www.prism-magazine.org/oct08/feature\\_03.cfm](http://www.prism-magazine.org/oct08/feature_03.cfm) accessed February 27, 2014).
- Lowrey, Annie. 2010. "Do We Need Google to Measure Inflation?" *Slate*. ([http://www.slate.com/articles/business/moneybox/2010/12/do\\_we\\_need\\_google\\_to\\_measure\\_inflation.html](http://www.slate.com/articles/business/moneybox/2010/12/do_we_need_google_to_measure_inflation.html) accessed May 7, 2014).
- McKinsey Global Institute. 2013. *Open data: Unlocking Innovation and Performance with Liquid Information*.
- MissionInsite. 2014. "Products Overview." (<http://churchinsite.com/products.html> accessed January 22, 2014).
- Messina, Joseph P., et al. 2006. "Evaluating Michigan's Community Hospital Access: Spatial Models for Decision Support." *International Journal of Health Geographics*. Vol. 5, p. 42. (<http://www.ij-healthgeographics.com/content/5/1/42> accessed July 7, 2014).
- Minnesota Population Center. 2014. "About MPC." (<https://www.pop.umn.edu/index.php> accessed May 8, 2014).
- Minority Business Development Agency. 2013. "Is There an Export Market for My Company?" ([http://www.mbda.gov/blogger/export-basics/there-export-market-my-company?utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=CWG\\_0816](http://www.mbda.gov/blogger/export-basics/there-export-market-my-company?utm_source=newsletter&utm_medium=email&utm_campaign=CWG_0816) accessed February 27, 2014).
- National Center for Education Statistics. 2013. Digest of Education Statistics. "Table 91. Number of Public School Districts and Public and Private Elementary and Secondary Schools: Selected Years, 1869-70." ([http://nces.ed.gov/programs/digest/d11/tables/dt11\\_091.asp](http://nces.ed.gov/programs/digest/d11/tables/dt11_091.asp))
- National Center for Education Statistics. 2013. "Projections of Education Statistics to 2021." (<http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2013008> accessed February 27, 2014).
- National Center for Education Statistics. 2014. 2012 Digest of Education Statistics. "Table 283. Degrees Conferred by Degree-Granting Institutions, by Level of Degree and Sex of Student: Selected Years, 1869-70 through 2020-21." ([http://nces.ed.gov/programs/digest/d11/tables/dt11\\_283.asp](http://nces.ed.gov/programs/digest/d11/tables/dt11_283.asp) accessed February 27, 2014).
- National Research Council. 2013. *Principles and Practices for a Federal Statistical Agency: Fifth Edition*. Washington, DC: The National Academies Press.
- National Research Council. 2013. *Benefits, Burdens, and Prospects of the American Community Survey: Summary of a Workshop*. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press, ([http://www.nap.edu/catalog.php?record\\_id=18259](http://www.nap.edu/catalog.php?record_id=18259) accessed July 7, 2014).
- Nielsen. 2013. "Nielsen Estimates 115.6 Million TV Homes in the U.S., Up 1.2%." (<http://www.nielsen.com/us/en/insights/news/2013/nielsen-estimates-115-6-million-tv-homes-in-the-u-s---up-1-2-.html> accessed July 7, 2014).
- Nielsen. 2014. 2013 *Nielsen Year in Review*. ([http://ir.nielsen.com/files/doc\\_financials/Nielsen\\_YIR\\_2013.pdf](http://ir.nielsen.com/files/doc_financials/Nielsen_YIR_2013.pdf) accessed June 2, 2014).
- Newburger, Eric. 2013. "Why the Census? Big Data from the Enlightenment to Today." Presentation at Wolfram Data Summit 2013 (September 5-6), Washington, DC.
- Office of Management and Budget. 1985. "Statistical Policy Directive Number 3" (50 Fed. Reg. 38,932-38,934 (September 25)).

- Office of Management and Budget. 2008. "Statistical Policy Directive Number 4" (73 Fed. Reg. 12,622-12,626 (March 7)).
- Office of Management and Budget. 2006. "Standards and Guidelines for Statistical Surveys." ([http://www.whitehouse.gov/sites/default/files/omb/assets/omb/inforeg/statpolicy/standards\\_stat\\_surveys.pdf](http://www.whitehouse.gov/sites/default/files/omb/assets/omb/inforeg/statpolicy/standards_stat_surveys.pdf) accessed July 7, 2014).
- Office of Management and Budget. 2014. Office of Information and Regulatory Affairs. (<http://www.reginfo.gov/public/do/eoCountsSearch> accessed January 22, 2014).
- Office of Management and Budget. 2013a. *Statistical Programs of the United States Government: Fiscal Year 2014*.
- Office of Management and Budget. 2013b. *The Budget: Historical Tables*. Table 5.1—Budget Authority by Function and Subfunction: 1976–2018 (<http://www.gpo.gov/fdsys/pkg/BUDGET-2014-TAB/xls/BUDGET-2014-TAB-5-1.xls> accessed February 21, 2014).
- Office of Management and Budget. 2013c. Office of Information and Regulatory Affairs. "Strengthening Federal Statistics" from the *FY 2014 Budget, Analytical Perspectives*, p. 345 (available at [http://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/strengthening\\_fed\\_stats\\_fy2014.pdf](http://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/strengthening_fed_stats_fy2014.pdf)).
- Open Data 500. 2014. (<http://www.opendata500.com/> accessed May 13, 2014).
- Oxera Consulting, Ltd. 2013. *What Is the Economic Impact of Geo Services?* (Funded by Google.)
- Parker, Robert P. and C. Brian Grove. 2000. "Census Bureau Moves Ahead on Measuring E-business." *Business Economics*. (July), pp. 63-65.
- Pierce, Justin. 2009. "Using Census Bureau Data for Real Estate Market Research." (<http://www.biggerpockets.com/renewsblog/2009/09/06/using-census-bureau-data-for-real-estate-market-research/> accessed July 7, 2014).
- Pillar, Matt. 2010. "PETCO Improves Location Selection." *Integrated Solutions For Retailers Magazine*, pp. 36-37.
- Police Foundation. 2000. "Crime Mapping News." (Fall 2000, Volume 2, no. 4.
- Reamer, Andrew. 2006. "To Take a Bite Out of Crime: Safeguard the Census." (<http://www.brookings.edu/research/opinions/2006/06/26communitydevelopment-reamer> accessed July 7, 2014).
- Robinson, David, Harlan Yu, William P. Zeller, and Edward W. Felten. 2009. "Government Data and the Invisible Hand." *Yale Journal of Law and Technology*. Vol. 11, pp. 160-175.
- Simmons, Tavia and Martin O'Connell. 2003. "Married-Couple and Unmarried-Partner Households: 2000." (<http://www.census.gov/prod/2003pubs/censr-5.pdf> accessed January 6, 2014).
- Specter, Michael. 2013. "Climate by the Numbers: Can a tech firm help farmers survive global warming?" *The New Yorker*. (November 11), pp. 38-43.
- Spoon River College. 2014. "Getting Ready for College." ([http://www.src.edu/admissions/high\\_school/Pages/getting-ready.aspx](http://www.src.edu/admissions/high_school/Pages/getting-ready.aspx) accessed February 27, 2014).
- Stiglitz, Joseph E., Peter R. Orszag, and Jonathan M. Orszag. October 2000. "The Role of Government in a Digital Age." Study commissioned by the Computer and Communications Industry Association. Washington, DC.

- Thompson Reuters. 2012 *Annual Report Regulatory Filing*. ([http://ar.thomsonreuters.com/2012/\\_files/pdf/2012-Annual-Report-Regulatory-Filing.pdf](http://ar.thomsonreuters.com/2012/_files/pdf/2012-Annual-Report-Regulatory-Filing.pdf) accessed June 3, 2014.)
- Trulia. 2014a. "What is Trulia?" (<https://trulia.zendesk.com/hc/en-us/articles/200053106-What-is-Trulia-> accessed January 3, 2014).
- Trulia. 2014b. "About Us: What we do." (<http://www.trulia.com/about> accessed January 3, 2014).
- Trulia. 2014c. "Bethesda Community Info." ([http://www.trulia.com/real\\_estate/Bethesda-Maryland/community-info/](http://www.trulia.com/real_estate/Bethesda-Maryland/community-info/) accessed January 3, 2014).
- U.S. Census Bureau. 2012. Video: "Target and the American Community Survey." ([http://www.youtube.com/watch?v=jgsdQxTv5kY&feature=player\\_embedded](http://www.youtube.com/watch?v=jgsdQxTv5kY&feature=player_embedded) accessed July 7, 2014).
- U.S. Census Bureau. 2013a. "Table 4: U.S. Retail Trade Sales - Total and E-commerce: 2011-1998." (<http://www.census.gov/econ/estats/2011/historical/2011ht4.xls> accessed January 21, 2014).
- U.S. Census Bureau. 2013b. "Table 4: December 2013 Construction at \$930.5 Billion Annual Rate". (<http://www.census.gov/construction/c30/pdf/pr201312.pdf> )
- U.S. Census Bureau. 2013d. *2012 Annual Services*. "Table 2: Estimated Revenue by Tax Status for Employer Firms: 2007 through 2012." (<http://www2.census.gov/services/sas/data/table2.xls> accessed January 3, 2014).
- U.S. Census Bureau. 2014a. "Small Area Income and Poverty Estimates." (<http://www.census.gov/did/www/saipe/> accessed January 6, 2014).
- United Nations Statistics Division. 2009. *System of National Accounts 2008*. (<http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf> accessed January 7, 2014).
- United Nations Statistics Division. 2014. "The System of National Accounts (SNA)." (<http://unstats.un.org/unsd/nationalaccount/sna.asp> accessed January 7, 2014).
- U.S. General Accountability Office. 2013. *Information Resellers: Consumer Privacy Framework Needs to Reflect Changes in Technology and the Marketplace*. (<http://www.gao.gov/products/GAO-13-663> accessed July 7, 2014).
- U.S. Office of the President. 2014. *Open Data Action Plan*. ([http://www.whitehouse.gov/sites/default/files/microsites/ostp/us\\_open\\_data\\_action\\_plan.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/us_open_data_action_plan.pdf) accessed July 7, 2014).
- Walters, Dan. 2013. "California Data Bank Could Fill Big Knowledge Gap." *The Sacramento Bee*. (<http://www.sacbee.com/2013/08/18/5659853/dan-walters-california-data-bank.html> accessed July 7, 2014).
- Yousuf, Hibah. 2013. "Thomson Reuters Ends Early Access to Key Market Data." *CNN Money*. (<http://money.cnn.com/2013/07/08/investing/thomson-reuters-consumer-sentiment/> accessed February 27, 2014).
- Zillow. 2014a. "What is Zillow?" (<http://www.zillow.com/corp/About.htm> accessed January 3, 2014).
- Zillow. 2014b. "What is a Zestimate?" (<http://www.zillow.com/wikipages/What-is-a-Zestimate/> accessed January 3, 2014).
- Zillow. 2014c. "Data Coverage and Zestimate® Accuracy." (<http://www.zillow.com/howto/DataCoverageZestimateAccuracy.htm> accessed January 3, 2014).
- Zillow. 2014d. "Bethesda Local Information." ([http://www.zillow.com/local-info/MD-Bethesda/r\\_37406/](http://www.zillow.com/local-info/MD-Bethesda/r_37406/) accessed January 3, 2014).

## Appendix I: Definitions of Terms Used

**Administrative Data** - Information collected and used by program, administrative, and regulatory agencies and not originally intended to be used for statistical purposes.

**Analysts** - GDIS firms that analyze Government data directly or indirectly, using third-party data products to create products for other firms.

**Benchmarks** - GDIS firms that create statistics or other data products using non-Government data, but use Government data as a reference and standard with which to adjust or weight or test the validity of their products.

**Data Brokers** - GDIS firms that compile Government data (and sometimes proprietary private sector data) from around the world and add value by aggregating, integrating, and facilitating access to and analysis of the data. These firms sell access to the resulting repackaged data and often provide analysis of the data and even news reporting on the data. In other words, these firms may be vertically integrated GDIS firms.

**Government Data** - A diverse set of Federal Government information and data products, including

- *Information published by Government agencies for statistical purposes.* Such information is compiled from raw data collected directly from individuals, households, businesses, other organizations, and governments. The information may also be indirectly compiled from Government administrative sources, private sector data sources, and other sources. The data may be summarized in tabulations by statistical agencies to describe specific characteristics of groups of persons, businesses, or other entities. The agencies make such information available for use by other Government agencies and the general public to inform decision making, in program analysis and evaluation, and research.
- *Information in Government administrative records on individuals and businesses.* Examples include publicly available data about private sector company-sponsored retirement plans reported to the Labor Department on Form 5500; Securities and Exchange Commission filings; Government contracts data; real estate appraisal records; and voter registration rolls. Historically, a Federal Freedom of Information Act request has often been required to make much of this category of information public, but recent “open data” efforts have aimed to make such information more easily accessible. However, much data remains strictly confidential, such as tax return data collected by the Internal Revenue Service or Social Security Administration data.
- *Physical measurements by Government agencies of natural phenomena.* Examples include weather and climate data such as local temperature readings collected by the National Weather Service of the National Oceanic and Atmospheric Administration Global Positioning System data; and measurements of the characteristics of elements and molecules collected and published by the National Institute for Standards and Technology Physical Measurement Laboratory.

**Government data-intensive sector (GDIS)** - The ecosystem of firms that rely heavily on Government data as a key input to their production processes for providing services to households, governments, and firms inside and outside of the GDIS. These firms consist of firms in NAICS industry codes defined in Appendix II.

**North American Industry Classification System (NAICS)** - The standard used by Federal statistical agencies in classifying business establishments for the purpose of collecting, analyzing, and publishing statistical data related to the U.S. business economy. Each industry is assigned a 6-digit numeric code; industry sectors at varying levels of aggregation consist of industries sharing the same first two or more digits in their NAICS codes (<http://www.census.gov/eos/www/naics/>).

**Principal Federal Statistical Agencies** - Thirteen agencies whose main activities are the collection, compilation, processing or analysis of information for statistical purposes. For most, funding appears as a separate line item in the President's Budget.

**Public Good** - A good or service that is non-rival and non-excludable.

**Value-Added Re-packagers** - GDIS firms that aggregate Government data from many different U.S. Government agencies, private-sector firms, other governments, and international organizations within a single application. The application adds value to the data by providing easy-to-use interfaces with the data and possibly the capability for users to create charts and tables and switch periodicity easily (for example, switch between monthly, quarterly, and annual series). Furthermore, there are synergies in using data from multiple sources, which is another way these companies add value. These companies often splice together time series for which the original source had a break in methodology, thus creating long time series that have value even if they do not meet rigorous statistical agency standards.

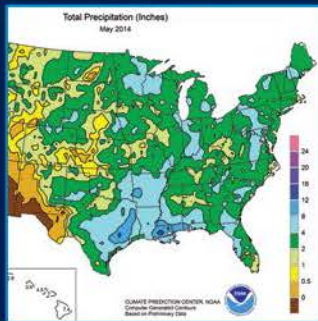


## Appendix II: Definitions of Government Data-Intensive Industries

NAICS Code	Industry Name	Industry Description
519110	News Syndicates	This industry comprises establishments primarily engaged in supplying information, such as news reports, articles, pictures, and features, to the news media.
519130	Internet Publishing and Broadcasting and Web Search	This industry comprises establishments primarily engaged in 1) publishing and/or broadcasting content on the Internet exclusively or 2) operating Web sites that use a search engine to generate and maintain extensive databases of Internet addresses and content in an easily searchable format (and known as Web search portals). The publishing and broadcasting establishments in this industry do not provide traditional (non-Internet) versions of the content that they publish or broadcast. They provide textual, audio, and/or video content of general or specific interest on the Internet exclusively. Establishments known as Web search portals often provide additional Internet services, such as e-mail, connections to other web sites, auctions, news, and other limited content, and serve as a home base for Internet users.
531390	Other Activities Related to Real Estate	This industry comprises establishments primarily engaged in performing real estate related services (except lessors of real estate, offices of real estate agents and brokers, real estate property managers, and offices of real estate appraisers).
541611	Administrative Management and General Management Consulting Services	This industry comprises establishments primarily engaged in providing operating advice and assistance to businesses and other organizations on administrative management issues, such as financial planning and budgeting, equity and asset management, records management, office planning, strategic and organizational planning, site selection, new business startup, and business process improvement. This industry also includes establishments of general management consultants that provide a full range of services, such as administrative; human resource; marketing; process, physical distribution, and logistics; or other management consulting services to clients.
541612	Human Resources Consulting Services	This industry comprises establishments primarily engaged in providing advice and assistance to businesses and other organizations in one or more of the following areas: (1) human resource and personnel policies, practices, and procedures; (2) employee benefits planning, communication, and administration; (3) compensation systems planning; and (4) wage and salary administration.

541613	Marketing Consulting Services	This industry comprises establishments primarily engaged in providing operating advice and assistance to businesses and other organizations on marketing issues, such as developing marketing objectives and policies, sales forecasting, new product developing and pricing, licensing and franchise planning, and marketing planning and strategy.
541614	Process, Physical Distribution, and Logistics Consulting Services	This industry comprises establishments primarily engaged in providing operating advice and assistance to businesses and other organizations in areas, such as: (1) manufacturing operations improvement; (2) productivity improvement; (3) production planning and control; (4) quality assurance and quality control; (5) inventory management; (6) distribution networks; (7) warehouse use, operations, and utilization; (8) transportation and shipment of goods and materials; and (9) materials management and handling.
541618	Other Management Consulting Services	This industry comprises establishments primarily engaged in providing management consulting services (except administrative and general management consulting; human resources consulting; marketing consulting; or process, physical distribution, and logistics consulting). Establishments providing telecommunications or utilities management consulting services are included in this industry.
541690	Other Scientific and Technical Consulting Services	This industry comprises establishments primarily engaged in providing advice and assistance to businesses and other organizations on scientific and technical issues (except environmental).
541910	Marketing Research and Public Opinion Polling	This industry comprises establishments primarily engaged in systematically gathering, recording, tabulating, and presenting marketing and public opinion data.





**U.S. DEPARTMENT OF COMMERCE  
ECONOMICS AND STATISTICS ADMINISTRATION**