The GovData Project
MIT-Harvard Winter Course 2011

Module 1:
Government Data: Its Power and Its Flaws
Q:
Why is Government data interesting?
Why is Government data interesting?

Bureau of Economic Analysis (BEA)

“The Bureau of Economic Analysis (BEA) is an agency of the United States Department of Commerce (DOC) that compiles data on a wide variety of important economic indicators.”

... like the GDP figures you always hear about.
Why is Government data interesting?

... “the commerce department reports that the GDP grew by 1.2% this quarter, prompting fears that ...”

That stuff refers to *this* dataset.
Why is Government data interesting?

Table 1.1.1. Percent Change From Preceding Period in Real Gross Domestic Product.

GDP volatility has decreased over time, even including most recent recession.
Why is Government data interesting?

Table 1.1.1. Percent Change From Preceding Period in Real Gross Domestic Product.

Durable goods are more volatile than nondurable goods, ... although this trend has attenuated recently.
Why is Government data interesting?

Table 1.1.5. Gross Domestic Product (absolute numbers)

Units: Billions of dollars
Why is Government data interesting?

Table 1.1.5. Gross Domestic Product (absolute numbers)

Durables roughly kept pace until early 1940s ...
Why is Government data interesting?

Table 1.1.5. Gross Domestic Product (absolute numbers)

Exports held on until Q1 1981, and then, wham ..
Defense / nondefense spending gap is at its highest, ever. (including WWII, Iraq II, &c)
Why is Government data interesting?

Table 1.1.10. Percentage Shares of Gross Domestic Product

Clearly a relative scale ...
Why is Government data interesting?

Table 1.1.10. Percentage Shares of Gross Domestic Product

Goods vs. services seems almost like a zero-sum-game ...

and the sides changed Q1 1970.
Why is Government data interesting?

Table 1.1.10. Percentage Shares of Gross Domestic Product

Exports vs. imports:
- until 1950s, exports > imports
- balanced until mid-70s
- significant account deficit thereafter

But its NOT a zero-sum game ... they’re coupled.
Module 0. Introduction

(needed background: None)
Intro: The Two Audiences

Data Professionals

“Front-end users.” (a.k.a. “the public”, “the American people” &c)
Intro: The Two Audiences

Data Professionals
- Academics
- Policy entrepreneurs
- Journalists
- Developers
- Data archives

“Front-end users.” (a.k.a. “the public”, “the American people” &c)
- Students
- Congressional staffers
- Newspaper readers (esp. online)
- Googlers (e.g. everyone)
- Academics & Policy entreprenuers
Intro: Huge Recent Hype

- **Open Government Initiative** | The White House
  Earlier this month, the White House Open Government Initiative and General Services Administration requested your input to help design an... [www.whitehouse.gov/open](http://www.whitehouse.gov/open)

- **Open Government Initiative Blog** | The White House
  Earlier this month, the White House Open Government Initiative and General... [www.whitehouse.gov/open/blog](http://www.whitehouse.gov/open/blog)

- **Open Government Initiative**
  Open Government Initiative - Transparency, public participation, and collaboration improvements will strengthen NARA’s crucial role in supporting democracy... [www.archives.gov/open](http://www.archives.gov/open)

- **Open Government Initiative**
  Date: 04/07/2010 Description: Open Government Plan - State Dept Image The Department of State has a longstanding history and commitment to sharing... [www.state.gov/open](http://www.state.gov/open)

- **Open Government Initiative**
  On December 8, 2009, the Office of Management and Budget (OMB) issued an Open Government Directive requiring Federal agencies to take specific actions to... [www.eeoc.gov/open](http://www.eeoc.gov/open)


- **Open Government Initiative**
  A Treasury Open Government Initiative: www.PRACComment.gov. Treasury launched the first government site to collect public comments on the Paperwork Reduction... [www.treasury.gov/open](http://www.treasury.gov/open)

- **White House launches open government initiative** - Nextgov
Our Theses:

1. Something important has been started.

2. There are some important glaring gaps that are not really being addressed.

3. You don’t need to be the US Federal Government to address them.

4. But you do need a substantial, organized, open source effort.
The point of the Govdata project

- to take as much tabular-format government data as possible,
- lightly structure it so it fits into a uniform DB format,
- hosted in a single location, that is updated regularly,
- with standardized temporal and spatial columns,
- and rich metadata;

and then:

- make the backend data accessible through powerful general use APIs with detailed query control
- apply cutting-edge indexing techniques to make individual data slices searchable
- integrate new javascript visualization techniques to design an effective and powerful frontend
Course Agenda:

- Give a brief history & overview of government data initiatives

- Explain the Big Gap, as we see it.

- Tell you a bit about our solution.

- Teach you parsing and visualization tools through:
  - boring lectures engaging presentations
  - extended hands-on coding sessions, to really bring it home

- Create the nucleus of an open source community.
Schedule

Day 1:  Government Data 101: It’s power & its flaws
The GovData Project: Principles
Web scraping & data cleaning, I

Day 2:  Web scraping & data cleaning, II
A short introduction to MongoDB & Apache Solr
The GovData APIs
Hands-On Coding Project Overview

Day 3:  Hands-on Coding Sessions I & II

Day 4:  Introduction to “Well Structured Javascript”
Hands-on Coding Sessions III

Day 5:  Hands-on Coding Sessions IV & V
The census tells us who we are and where we are going as a nation. The census helps our communities determine where to build everything from schools to superhighways, and to draw the lines of legislative districts and reapportion the seats each State holds in Congress.

1990 Census of Population

CP-1 - General Population Characteristics

CP-2 - Social and Economic Characteristics

CP-3-1 - The Foreign-Born Population in the United States [PDF, 9.3MB]

CP-3-2 - Ancestry of the Population in the United States [PDF, 13.4MB]

CP-3-3 - Persons of Hispanic Origin in the United States [PDF, 6.7MB]

CP-3-4 - Education in the United States [PDF, 9.5MB]

CP-3-5 - Asians and Pacific Islanders in the United States [PDF, 6.3MB]

CP-3-6 - Characteristics of the Black Population [PDF, 8.2MB]

CP-3-7-1 - Characteristics of American Indians by Tribe and Language: Section 1 [PDF, 14.4MB]

CP-3-7-2 - Characteristics of American Indians by Tribe and Language: Section 2 [PDF, 10.7MB]

CP-S-1-1 - Detailed Occupation and Other Characteristics From the EFO File for the United States [PDF, 2.6MB]

CP-S-1-2 - Detailed Ancestry Groups for States [PDF, 2.5MB]
Collecting Data since ...

U.S. Census Bureau

CENSUS OF POPULATION AND HOUSING
1980 Census

For most of the United States, "Census Day" for the 1980 enumeration was April 1, 1980.\(^\d\) As in past censuses, all questionnaires were to be c

The 1980 census also included two small surveys: the Components of Inventory Change Survey, which obtained information on counts and s to supern

1990 Census of Population

CP-1 - 1980 Census of Population

CP-2 - Vol. I. Characteristics of the Population

CP-3-1 Vol. II. Subject Reports

CP-3-2 Supplementary Reports

CP-3-3 1980 Census of Housing

CP-3-4

CP-3-5 NOTE: Currently, none of the 1980 Census of Housing reports are available online. For a listing of reports that were pu

CP-3-6 For access to volumes not available, please contact your local Federal Depository Library.

CP-3-7 1980 Census of Population and Housing

CP-3-7. Preliminary Population and Housing Unit Counts. (PHC80-P-1 to -56). (unavailable)

CP-S-1. Final Population and Housing Unit Counts. (PHC80-V). (unavailable)

CP-S-1. Block Statistics. [Microfiche]. (unavailable)
Collecting Data since ...

U.S. Census Bureau

CENSUS OF POPULATION AND HOUSING

1970 Census

1990

The 1980 Survey

For most

1990

When planning for the 1970 census, the need for an accurate count of the population was even greater than in the past because of the increasing tendency for government officials and others of the potential effects of census undercounts. [More/Less]

CP-1 - (1980)

1970 Census of Population

CP-2 - Vol. I

 Characteristics of the Population

CP-3-1 Vol. II

 Series PC(S1). Supplementary Reports

Series PC(S2). Subject Reports

CP-3-2 Supplier

 Advance Reports

CP-3-3 1980

 Series PC(V1). Final Population Counts. * - PDF, ZIP [138.3 MB]

Series PC(V2). General Population Characteristics. * - PDF, ZIP [115.6 MB]

CP-3-4

NOTE:

CP-3-5 Preliminary Reports

For acc...

CP-3-6 1980

 Series PC(P1). Population Counts for States – Preliminary Reports * - PDF, ZIP [26.1 MB]

Series PC(P2). Population Counts for Standard Metropolitan Statistical Areas – Preliminary Reports * - PDF, ZIP [84.0 MB]

CP-3-7 Preliminary

Population of Congressional Districts.

Population of Cities of 25,000 Inhabitants or More.

Population of Standard Metropolitan Statistical Areas.

CP-S-1. Final P.

CP-S-1. Block S Other Reports * - PDF, ZIP [91.7 MB]

Alphabetical Index of Industries and Occupations.
Collecting Data since ...

U.S. Census Bureau

The census assistance

For most

The 1986 Survey, in

When p

local go

1990

CP-1 - - 1980

CP-2 - - Vol. I.

CP-3-1 Vol. II.

CP-3-2 Suppl.

CP-3-3 Advan.

CP-3-4 Volume 1: Population, 1920. Number and distribution of inhabitants, 695 pp., plus one 4-color foldout map. * - PDF, ZIP [154.9 MB]


CP-3-6 Volume 3. Population, 1920. Composition and characteristics of the population by states. 1253 pp., plus one 4-color foldout map

NOTE: Preliminary


CP-3-8 Volume 5. Agriculture. General report and analytical tables. 935 pp * - PDF, ZIP [195.2 MB]


Part 2. The Southern states. 746 pp. (unavailable)

Part 3. The Western states and outlying possessions. 423 pp. (unavailable)

CP-3-7 Preliminary

CP-3-8

CP-3-9

Volume 7. Irrigation and drainage. General report and analytical tables and reports for states, with statistics for counties. 741 pp.,
Collecting Data since ...

U.S. Census Bureau

CENSUS OF POPULATION AND HOUSING
1790 Census

The first enumeration began on Monday, August 2, 1790, little more than a year after the inauguration of President Washington, and under the U.S. judicial districts under an act that, with minor modifications and extensions, governed census-taking through 1900. The result was that, there to remain for the inspection of all concerned...” and that “the aggregate amount of each description of persons in each household of the following descriptions: Free White males of 16 years and upward (to assess the country's wealth and population for taxation and representation)."

1790 Census Information

Return of the whole number of persons within the several districts of the United States...

Heads of Families at the First Census of the United States in the year 1790

Connecticut - PDF, ZIP [60.6 MB]
Maine - PDF, ZIP [20.3 MB]
Maryland - PDF, ZIP [38.2 MB]
Massachusetts - PDF, ZIP [81.4 MB]
New Hampshire - PDF, ZIP [30.5 MB]
New York - PDF, ZIP [74.0 MB]
North Carolina - PDF, ZIP [107.4 MB]
Pennsylvania - PDF, ZIP [94.8 MB]
Rhode Island - PDF, ZIP [14.3 MB]
South Carolina - PDF, ZIP [54.2 MB]
Vermont - PDF, ZIP [21.9 MB]
Virginia - PDF, ZIP [67.7 MB]

For online viewing, each book with a total file size over 10 MB has been divided into several internally linked PDF files. The total file size of the online PDF files is similar to that of their ZIP counterparts.

Users with visual impairments who have difficulty accessing PDF documents may access technical support or call 301-

“... 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.”

-- Art. I, Sec. II, US Constitution
FedStats in late 1990s

FedStats celebrating over 10 years of making statistics from more than 100 agencies available to citizens everywhere.

**Links to statistics**

- **Topic Links - A To Z** - Direct access to statistical data on topics of your choice.
- **MapStats** - Statistical profiles of States, counties, cities, Congressional Districts, and Federal judicial districts.
- **Statistics By Geography From U.S. Agencies** - International comparisons, national, State, county, and local.
- **Statistical Reference Shelf** - Published collections of statistics available online including the Statistical Abstract of the United States.
- **Search** across agency websites.

**Links to statistical agencies**

- **Agencies Listed Alphabetically** - with descriptions of the statistics they provide and links to their websites, contact information, and key statistics.
- **Agencies by subject** - Select a subject:
  - Agriculture
  - Submit
- **Press Releases** - The latest news and announcements from individual agencies.
- **Kids' Pages** - on agency websites.
- **Data Access Tools** - Selected agency online databases.

+ uncountably many other comparatively low-profile initiatives ...
MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

SUBJECT: Transparency and Open Government

My Administration is committed to creating an unprecedented level of openness in Government. We will work together to ensure the public trust and establish a system of transparency, public participation, and collaboration. Openness will strengthen our democracy and promote efficiency and effectiveness in Government.
EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

December 8, 2009

M-10-06

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Peter R. Orszag
        Director

SUBJECT: Open Government Directive

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

The three principles of transparency, participation, and collaboration form the
The Open Government Initiative

Due to efforts and passions at various offices in the federal government, there has been an explosion of interest in “open data.”

Vivek Kundra
Federal CIO

Aneesh Chopra
Federal CTO
The Open Government Initiative

Due to efforts and passions at various offices in the federal government, there has been an explosion of interest in “open data.”

Holy alliance of lawyers, economists, and programmers ...
The Open Government Initiative

CIO Council Members

By Department

Department of Agriculture

Mr. Chris Smith
Chief Information Officer
1400 Independence Avenue, SW
Room 414W
Washington, DC 20250
chris.smith@usa.gov
Office Phone: 202-720-8833
Fax: 202-720-1031

Mr. Charles McClan
Deputy Chief Information Officer
1400 Independence Avenue
432W Whitney Building
Washington, DC 20416
charles.mcclan@usda.gov
Office Phone: 202-720-8865
Fax: 202-720-3733

Department of the Air Force

Lt. Gen. William T. Lord
Chief of Warfighting Integration and Chief Information Officer
1800 Air Force Pentagon
Room 4E1050
Washington, DC 20330-1800
subxc.workflow@pentagon.mil
Office Phone: 703-695-6829
Fax: 703-695-7812

Mr. Daniel F. McMillin
Deputy Chief of Warfighting Integration and Deputy Chief Information Officer
1800 Air Force Pentagon
Room 4E1050
Washington, DC 20330-1800
daniel.mcmillin@pentagon.mil
Office Phone: 703-697-1605
Fax: 703-692-7812

Department of Health & Human Services

Mr. Michael W. Carleton
Chief Information Officer
200 Independence Avenue, SW
Hubert Humphrey Building, Room 537H
Washington, DC 20201
michael.carleton@hhs.gov
Office Phone: 202-690-5162
Fax: 202-690-5715

Mr. John Teeter
Deputy Chief Information Officer
200 Independence Avenue, SW
Hubert Humphrey Building, Room 537H
Washington, DC 20201
john.teeter@hhs.gov
Office Phone: 202-690-5162
Fax: 202-690-5715

Department of Homeland Security

Mr. Richard Spies
Chief Information Officer
301 7th Street
Washington, DC 20528
cio@dhs.gov
Office Phone: 202-447-3735

Ms. Margie Graves
Deputy Chief Information Officer
Department of Homeland Security
301 7th Street
Washington, DC 20528
cio@dhs.gov
Office Phone: 202-447-3735
The Open Government Initiative

data.gov
data.gov.uk
datasf.org
data.dc.gov
Mass ODI
NYC Datamine
Toronto.ca/open
socrata.com
sunlight labs
The Open Government Initiative

data.gov ---> the professionals ---> front-end users
The Open Government Initiative

data.gov.uk

Opening up government

Data

Looking for something specific, or just want to know more about how Government spends your money? You'll find 5,300 datasets to help you get answers.

Browse by publisher

- Advisory, Conciliation and Arbitration Service (0)
- Arts and Humanities Research Council (0)
- Arts Council England (1)
- Ashfield District Council (1)
- Ashton, Leigh and Wigan Community Healthcare NHS Trust (0)
- Attorney General's Office (3)
- Audit Commission (0)

Browse by nation

- England
- Northern Ireland
- Scotland
- Wales

Popular tags

- health (1,988)
- care (1,324)
- communities (1,306)
- child (1,130)
- health-and-social-care (1,104)
- transparency (1,062)
- children (1,058)
- local-government (1,058)
The Open Government Initiative

DataSF is a clearinghouse of datasets available from the City & County of San Francisco. Our goal in releasing this site is:
(1) improve access to data
(2) help our community create innovative apps
(3) understand what datasets you'd like to see
(4) get feedback on the quality of our datasets.

Datasets

**Film Locations in San Francisco**
Categorized under Admin & Finance Views: 1123 (http://www.filmsf.org)

Description: If you love movies, and you love San Francisco, you're bound to love this -- a listing of filming locations of movies shot in San Francisco from 1924 - 2010. You'll find the titles, locations, fun facts, names of the director, writer, actors, and studio for most of these films.
Agency Name: San Francisco Film Commission
Date Released: 12/21/2010
Time Period: 1924 - 2010
Frequency: Annual
Location of dataset: Film Locations in SF Dataset
Format: excel
Location of Data Dictionary: www.filmsf.org

Tag cloud

**San Francisco Wind Monitoring Data**
Categorized under Environment Views: 275 (Editorial)
The Open Government Initiative

Sunlight LABS
We're developers & designers opening America's government

National Data Catalog

This project is funded by Sunlight | Became a project on August 12, 2009

The National Data Catalog aims to be a complete catalog of all data sets and APIs that are either put out by the government or are derived from the government. Scoped to all government levels (federal, state, and local), and all branches (executive, legislative, judicial), NDC will be the one-stop shop for developers, researchers, and investigative journalists interested in government data.

NDC will tap into the social benefits of having users come together around common interests. More than just a catalog, it will be a place for community-supported documentation about government data.

Tags

data
ruby

Project Participants

Lead
David James

Team
Luigi Montaner

Join This Project?
If you're interested in helping out on
The Open Government Initiative

Code for America Fellowship PSA: "What if...?"

Tim O'Reilly
Founder and CEO, O'Reilly Media

Code for America was founded to help the brightest minds of the Web 2.0 generation transform city governments. Cities are under greater pressure than ever, struggling with budget cuts and outdated technology. What if, instead of cutting services or raising taxes, cities could leverage the power of the web to become more efficient, transparent, and participatory?

We believe there is a wealth of talent in the web industry eager to contribute to the rebuilding of America. Code for America gives them the means.

City governments, learn more.
The Open Government Initiative

web APIs
web catalogs

API = “application programming interface”
= URL access scheme, in web context
http://mysite.com/?title=MyLife&author=Me

NYTimes

NYC datamine

sunlight labs

codefor america .org

data.gov

sfdata .org

apps

websites

ppl

ppl

ppl

ppl

ppl

ppl

ppl

ppl

ppl

ppl

ppl

ppl

ppl

ppl

api = organized metadata, searchable

Apps = cool iPad stuff
The Datasets Themselves

There used to be this great page on data.gov ...

Listing all the participating agencies -- that is, all the actual data sources.

Unfortunately, it’s no longer accessible.

---

Hi Dan,

No problem! Glad I could help. Keep me in the loop about how things progress with NTIA.

Not sure where the page you’re looking for went – I did a quick search myself and didn’t see anything…sorry!

Best,

[Redacted]
The Datasets Themselves

Let’s explore it a bit ..
The Datasets Themselves

Data has a natural **source tree hierarchy:**

Agency > Subagency > Program > Topic . . . Dataset

Depending on how you count .... 2 ~ 10 TB
Claim:
The existing situation has a fundamental flaw.
The Fundamental Flaw

Catalogs:

NOT

Data collections:
The Fundamental Flaw

Catalogs: centralized formatting/storage of supplied metadata

NOT

Data collections: centralized storage of the data itself
The Fundamental Flaw

Employment, Hours, and Earnings from the Current Employment Statistics survey (National)

**DATASET SUMMARY**

- **Agency:** Department of Labor
- **Sub-Agency:** US Bureau of Labor Statistics
- **Category:** Labor Force, Employment, and Earnings
- **Date Released:** 1909
- **Date Updated:** Monthly
- **Time Period:** Employment and wages for the pay period including the 12th of the month
- **Frequency:** Monthly
- **Description:** Each month the Current Employment Statistics (CES) program surveys about 150,000 businesses and government agencies, representing approximately 390,000 individual worksites, in order to provide detailed industry data on employment, hours, and earnings of workers on nonfarm payrolls.

**DATASET RATINGS**

- **Overall:** ★★★★★
  - **Current:** (0 votes)
  - **Your Rating:** ★★★★★
- **Data Utility:** ★★★★★★
  - **Current:** (0 votes)
  - **Your Rating:** ★★★★★
- **Usefulness:** ★★★★★☆
  - **Current:** (0 votes)
  - **Your Rating:** ★★★★★[
- **Ease of Access:** ★★★★★
  - **Current:** (0 votes)
  - **Your Rating:** ★★★★★

**DOWNLOAD INFORMATION**

- **XML**
- **CSV/TXT**
- **XLS**
- **KML/KMZ**
- **Shapefile**
- **Maps**

**Dataset Information**

*Comment on these data:*
The Fundamental Flaw
The Fundamental Flaw

Interactive Access To
National Income and Product Accounts Tables

- Choose one of the Frequently Requested NIPA Tables
- Choose a table from a list of Selected NIPA Tables
- Choose a table from a list of All NIPA Tables
- Choose a table from the Keyword Index

What's New

February 1, 2010
Monthly NIPA tables have been updated to include the release of December 2010.

January 29, 2010
Selected Tables have been updated to include the advance estimate of the 4th quarter.

December 23, 2009
Monthly NIPA tables have been updated to include the release of November 2009.

December 22, 2009
Selected Tables have been updated to include the third estimate of the 3rd quarter.

November 25, 2009
Monthly NIPA tables have been updated to include the release of October 2009.
The Fundamental Flaw

All NIPA Tables

1 - Domestic Product and Income
2 - Personal Income and Outlays
3 - Government Current Receipts and Expenditures
4 - Foreign Transactions

Data Availability Legend: (A) Annually (Q) Quarterly (M) Monthly

Section 1 - Domestic Product and Income

Table 1.1.1. Percent Change From Preceding Period in Real Gross Domestic Product (A) (Q)
Table 1.1.2. Contributions to Percent Change in Real Gross Domestic Product (A) (Q)
Table 1.1.3. Real Gross Domestic Product, Quantity Indexes (A) (Q)
Table 1.1.4. Price Indexes for Gross Domestic Product (A) (Q)
Table 1.1.5. Gross Domestic Product (A) (Q)
Table 1.1.6. Real Gross Domestic Product, Chained Dollars (A) (Q)
Table 1.1.6A. Real Gross Domestic Product, Chained (1937) Dollars (A)
Table 1.1.6B. Real Gross Domestic Product, Chained (1952) Dollars (A) (Q)
Table 1.1.6C. Real Gross Domestic Product, Chained (1972) Dollars (A) (Q)
Table 1.1.6D. Real Gross Domestic Product, Chained (1987) Dollars (A) (Q)
Table 1.1.7. Percent Change From Preceding Period in Prices for Gross Domestic Product (A) (Q)
Table 1.1.8. Contributions to Percent Change in the Gross Domestic Product Price Index (A) (Q)
Table 1.1.9. Implicit Price Deflators for Gross Domestic Product (A) (Q)
Table 1.1.10. Percentage Shares of Gross Domestic Product (A) (Q)
Table 1.2.1. Percent Change From Preceding Period in Real Gross Domestic Product by Major Type of Product
Table 1.2.2. Contributions to Percent Change in Real Gross Domestic Product by Major Type of Product
Table 1.2.3. Real Gross Domestic Product by Major Type of Product, Quantity Indexes (A) (Q)
Table 1.2.4. Price Indexes for Gross Domestic Product by Major Type of Product (A) (Q)
Table 1.2.5. Gross Domestic Product by Major Type of Product (A) (Q)
Table 1.2.6. Real Gross Domestic Product by Major Type of Product, Chained Dollars (A) (Q)
Table 1.3.1. Percent Change From Preceding Period in Real Gross Value Added by Sector (A) (Q)
Table 1.3.3. Real Gross Value Added by Sector, Quantity Indexes (A) (Q)

Monthly NIPA tables have been updated to include the release of October 2009.
### National Income and Product Accounts Table

#### Table 1.1.4. Price Indexes for Gross Domestic Product

[Index numbers, 2005=100]  Seasonally adjusted

<table>
<thead>
<tr>
<th>First Year</th>
<th>Last Year</th>
<th>Series:</th>
<th>Data Table Options</th>
<th>Tools Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>1947-48 A &amp; Q</td>
<td>2009-10 A &amp; Q</td>
<td>Annual(A)</td>
<td><a href="#">HTML</a> Locking Stubs</td>
<td>Get all years Update</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Durable goods</td>
<td>41.332</td>
<td>41.712</td>
<td>41.966</td>
<td>42.493</td>
<td>42.818</td>
<td>43.377</td>
<td>44.886</td>
<td>44.967</td>
</tr>
<tr>
<td>Equipment and software</td>
<td>33.467</td>
<td>34.877</td>
<td>35.661</td>
<td>35.978</td>
<td>36.584</td>
<td>38.375</td>
<td>39.066</td>
<td>38.988</td>
</tr>
<tr>
<td>Change in private inventories</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Net exports of goods and services</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Services</td>
<td>16.207</td>
<td>15.503</td>
<td>15.142</td>
<td>15.020</td>
<td>15.141</td>
<td>15.234</td>
<td>15.357</td>
<td>15.468</td>
</tr>
<tr>
<td>Goods</td>
<td>4.840</td>
<td>5.072</td>
<td>5.291</td>
<td>5.229</td>
<td>5.248</td>
<td>5.140</td>
<td>5.052</td>
<td>4.945</td>
</tr>
</tbody>
</table>

*Note: The table above represents price indexes for Gross Domestic Product.*
### Table 1.1.4. Price Indexes for Gross Domestic Product

[Index numbers, 2005=100]

Seasonally adjusted

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Durable goods</td>
<td>41.332</td>
<td>41.712</td>
<td>41.945</td>
<td>42.491</td>
<td>42.815</td>
<td>43.377</td>
<td>44.886</td>
<td>46.406</td>
</tr>
<tr>
<td>10</td>
<td>Equipment</td>
<td>33.467</td>
<td>34.877</td>
<td>35.661</td>
<td>35.978</td>
<td>36.843</td>
<td>38.375</td>
<td>38.962</td>
<td>39.231</td>
</tr>
<tr>
<td>12</td>
<td>Change in private inventories</td>
<td>34.125</td>
<td>34.877</td>
<td>35.661</td>
<td>35.978</td>
<td>36.843</td>
<td>38.375</td>
<td>38.962</td>
<td>39.231</td>
</tr>
<tr>
<td>15</td>
<td>Services</td>
<td>16.207</td>
<td>15.503</td>
<td>15.142</td>
<td>15.02</td>
<td>15.141</td>
<td>15.234</td>
<td>15.357</td>
<td>15.217</td>
</tr>
<tr>
<td>16</td>
<td>Imports</td>
<td>14.707</td>
<td>15.742</td>
<td>16.585</td>
<td>17.151</td>
<td>17.551</td>
<td>17.611</td>
<td>17.437</td>
<td>17.244</td>
</tr>
<tr>
<td>23</td>
<td>State and loc</td>
<td>7.233</td>
<td>7.357</td>
<td>7.515</td>
<td>7.759</td>
<td>8.039</td>
<td>8.206</td>
<td>8.503</td>
<td>8.503</td>
</tr>
</tbody>
</table>
The Fundamental Flaw

Datasets

**Film Locations in San Francisco**
- Categorized under: Admin & Finance
- Views: 1123 (http://www.filmsf.org)

**Description:** If you love movies, and you love San Francisco, you're bound to love this -- a listing of filming locations of movies shot in San Francisco from 1924 - 2010. You'll find the titles, locations, fun facts, names of the director, writer, actors, and studio for most of these films.

**Agency Name:** San Francisco Film Commission

**Date Released:** 12/21/2010

**Time Period:** 1924 - 2010

**Frequency:** Annual

**Location of Dataset:** Film Locations in SF Dataset

**Format:** excel

**Location of Data Dictionary:** www.filmsf.org

Read More | 1 Comment | Share This | Tags: film locations movies actors directors

**San Francisco Wind Monitoring Data**
- Categorized under: Environment
- Views: 275 (Editorial)

---

What is DataSF?

DataSF is a clearinghouse of datasets available from the City & County of San Francisco. Our goal in releasing this site is:

1. Improve access to data
2. Help our community create innovative apps
3. Understand what datasets you'd like to see
4. Get feedback on the quality of our datasets.

Looking for a dataset or app that you'd like to see?

Tell us
The Fundamental Flaw

- to take as much tabular-format government data as possible,
- lightly structure it so it fits into a uniform DB format,
- hosted in a single location, that is updated regularly,
- with standardized temporal and spatial columns,
- and rich metadata;

and then:

- make the backend data accessible through powerful general use APIs with detailed query control
- apply cutting-edge indexing techniques to make individual data slices searchable
- integrate new javascript visualization techniques to design an effective and powerful frontend
The Fundamental Flaw

- to take as much tabular-format government data as possible,
- lightly structure it so it fits into a uniform DB format,
- hosted in a single location, that is updated regularly,
- with standardized temporal and spatial columns,
- and rich metadata;

and then:

- make the backend data accessible through powerful general use APIs with detailed query control
- apply cutting-edge indexing techniques to make individual data slices searchable
- integrate new javascript visualization techniques to design an effective and powerful frontend
The Fundamental Flaw

- to take as much tabular-format government data as possible,
- lightly structure it so it fits into a uniform DB format,
- hosted in a single location, that is updated regularly,
- with standardized temporal and spatial columns,
- and rich metadata;
and then:

- make the backend data accessible through powerful general use APIs with detailed query control
- apply cutting-edge indexing techniques to make individual data slices searchable
- integrate new javascript visualization techniques to design an effective and powerful frontend
The Fundamental Flaw

- to take as much tabular-format government data as possible,
- lightly structure it so it fits into a uniform DB format,
- hosted in a single location, that is updated regularly,
- with standardized temporal and spatial columns,
- and rich metadata;

and then:

- make the backend data accessible through powerful general use APIs with detailed query control
- apply cutting-edge indexing techniques to make individual data slices searchable
- integrate new javascript visualization techniques to design an effective and powerful frontend
The Fundamental Flaw

to take as much tabular-format government data as possible,
lightly structure it so it fits into a uniform DB format,
hosted in a single location, that is updated regularly,
with standardized temporal and spatial columns,
and rich metadata;
and then:
make the backend data accessible through powerful general use APIs with detailed query control
apply cutting-edge indexing techniques to make individual data slices searchable
integrate new javascript visualization techniques to design an effective and powerful frontend
EPA CTO told us:
Open Data Initiative NEVER intends to have centralized storage

Each dep’t is encouraged to choose its own solution. Each dataset its own API. (when there is one)

Initiative is not funded.

And ...

Sunlight Labs projects do not envision doing it either ...

Although ...

DataWiki Project (at RPI)
The Fundamental Flaw

One possible solution is to give developers access to lots of “third party tools” for data cleaning.
The Fundamental Flaw

All Data Collections

1995 Survey of Adults on Probation (SAP)
The 1995 Survey of Adults on Probation was the first nationally representative survey of probationers through a review of probationers' administrative records, and collected data on the frequency of parole revocation hearings.

2006 Census of State Parole Supervising Agencies
The 2006 Census of State Parole Supervising Agencies collected data from parole agencies in the U.S. on an annual basis, beginning and ending each year, the number of adults entering and exiting parole supervision, and parole supervision service availability.

Annual Probation Survey and Annual Parole Survey
Collect data from probation and parole agencies in the U.S. on an annual basis, beginning and ending each year, the number of adults entering and exiting parole supervision, and parole supervision service availability.

Annual Survey of Jails
Collect data from a nationally representative sample of local jails on jail inmates.

Capital Punishment (NPS-8)
Provides an annual summary of executions and executions that were stopped.

Census of Adult Parole Supervising Agencies
The 2006 Census of Adult Parole Supervising Agencies collected information on parole supervision, including the number of parole supervisees and the number of parolee supervision appointments.

Census of Federal Law Enforcement Officers
Collects data on all federal law enforcement agencies with arrest and investigation activities.

Census of Jail Inmates (2005)
Conducted approximately every five to seven years. Based on a complete enumeration of inmates, inmate counts, and movements, and persons supervised in the community.

Census of Jails
Conducted approximately every five to seven years. Provides information on individual jails.

Census of Law Enforcement Aviation Units (CLEAU)
Collects data on all law enforcement agencies with 100 or more sworn officers, including their locations, available assets and range, personnel, expenditures, and functions, and each could provide the needs of a nation emergency.

Census of Law Enforcement Training Academies
Collects data on the number and types of training academies, including basic organizational data, the survey collects information on training facilities.

Bureau of Justice Statistics
Filename: opspa06.csv
Report title: Characteristics of State Parole Supervising Agencies, 2006 NCJ 222180
Data source: Census of Adult Parole Supervising Agencies, CJ-36
Author(s): Thomas P. Bonczar
Refer questions to: asbjs@.usdoj.gov (202)307-0765
Date of version: 2/10/09
Appendix table 1. Persons under supervision by adult parole supervising agencies, 6/30/2006
Region and jurisdiction | Persons under supervision | Persons under adult parole supervision | Adult probation | State total | Supervised juvenile |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>137,250</td>
<td>151,113</td>
<td>5</td>
<td>35,673</td>
<td>2</td>
</tr>
<tr>
<td>Connecticut</td>
<td>2,239</td>
<td>2,998</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maine</td>
<td>8,018</td>
<td>32</td>
<td>yes</td>
<td>7,966</td>
<td>yes</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>3,382</td>
<td>3,362</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Hampshire</td>
<td>6,346</td>
<td>1,572</td>
<td>yes</td>
<td>4,674</td>
<td>yes</td>
</tr>
<tr>
<td>New Jersey</td>
<td>10,770</td>
<td>13,770</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td>53,379</td>
<td>53,213</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>26,733</td>
<td>24,966</td>
<td>yes</td>
<td>3,777</td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>11,770</td>
<td>512</td>
<td>yes</td>
<td>11,267</td>
<td></td>
</tr>
<tr>
<td>Vermont</td>
<td>9,025</td>
<td>1,050</td>
<td>yes</td>
<td>7,969</td>
<td></td>
</tr>
<tr>
<td>Midwest</td>
<td>461,007</td>
<td>123,370</td>
<td>7</td>
<td>325,087</td>
<td>5</td>
</tr>
<tr>
<td>Illinois</td>
<td>33,354</td>
<td>33,354</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indiana</td>
<td>8,481</td>
<td>7,863</td>
<td>yes</td>
<td>659</td>
<td></td>
</tr>
<tr>
<td>Iowa</td>
<td>26,291</td>
<td>3,973</td>
<td>yes</td>
<td>22,318</td>
<td></td>
</tr>
</tbody>
</table>

Bureau of Justice Statistics
Filename: dutt01.csv
Report title: Drug Use, Testing, and Treatment in Jails, NCJ-179999
Author: Doris James Wilson
Refer questions to: asbjs@usdoj.gov (202)307-0765
Date of version: 05/10/00
Table 1. Jail jurisdictions reporting drug testing policies, by size of jurisdiction, 1998

<table>
<thead>
<tr>
<th>Size of jurisdiction</th>
<th>Jurisdictions</th>
<th>Percent of jurisdictions with testing policies</th>
<th>Inmates</th>
<th>Staff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fewer than 50</td>
<td>1482</td>
<td>81.1 %</td>
<td>43.5 %</td>
<td>37.5 %</td>
</tr>
<tr>
<td>50-99</td>
<td>519</td>
<td>80.4</td>
<td>57.7</td>
<td></td>
</tr>
<tr>
<td>100-249</td>
<td>473</td>
<td>79.0</td>
<td>50.2</td>
<td></td>
</tr>
<tr>
<td>250-499</td>
<td>188</td>
<td>86.1</td>
<td>65.9</td>
<td></td>
</tr>
<tr>
<td>500-999</td>
<td>125</td>
<td>87.0</td>
<td>62.6</td>
<td></td>
</tr>
<tr>
<td>1,000 or more</td>
<td>123</td>
<td>92.0</td>
<td>66.6</td>
<td></td>
</tr>
</tbody>
</table>
The Fundamental Flaw
The Fundamental Flaw

... it’s not just a matter of “cleaning” individual files.

If everyone cleans it up in a slightly different way, it’s almost as bad as no cleaning at all.

There is a fundamental lack of “local integration.”

There needs to be a place (at least one) where such things are done in a centralized fashion.
Our realization

There needs to be a place (at least one) where all the data is hosted, and local integration is done in a centralized fashion.

I) it’s totally feasible, with the right tools

II) it’s not *that* expensive

III) for vast majority of data, no permission needed

IV) but it requires really good tools, and a focused community.