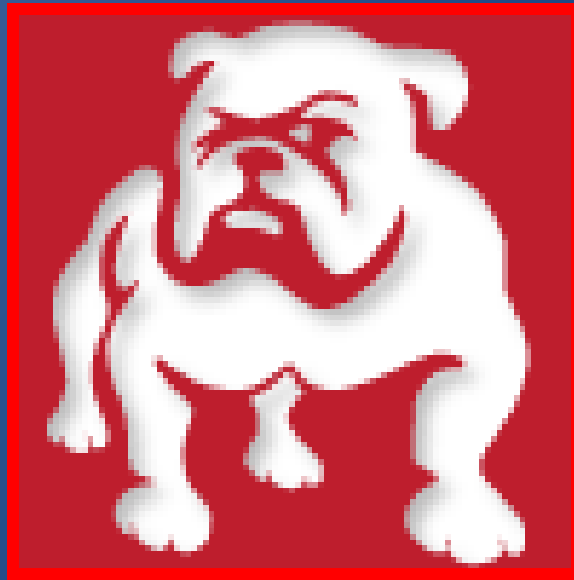


# Finding data and documents online



Jaimi Dowdell  
IRE/NICAR

# Where to look?

- ⦿ Data and documents already available:
  - Federal, state and local
  - Searchable databases
- ⦿ Tips for hunting down not-so-readily-available data.

# Things to keep in mind:

- ⦿ There is no magic site.
- ⦿ Half the battle is knowing what exists then figuring out how to search.
- ⦿ Never hesitate to request information you find in a more up-to-date format.
- ⦿ Some searchable databases do not include the entire dataset.
- ⦿ Clues for finding data are everywhere.

# Basic documents for any beat

(Tip: get more than one year)

- ⦿ Annual reports
- ⦿ Audits
- ⦿ Budgets
- ⦿ Salaries
- ⦿ Phone directories
- ⦿ Organizational charts
- ⦿ Vendors
- ⦿ Calendars for officials
- ⦿ Travel records

# Retention schedules

- ⦿ A plan to keep, maintain and dispose of records.
- ⦿ Treat the retention schedule like a menu of documents
  - Kentucky
  - List from the Council of State Archivists

# Take a tour

- ⦿ Visit the agency or office when you aren't working on a story
  - ▣ Grab every blank form.
  - ▣ Ask about records kept.
  - ▣ Ask how they do their jobs.
- ⦿ Meet document and data clerks and the IT folks.
- ⦿ Tour the website.

# Other ways to find what they keep...

- FOIA – Open records logs (example: 2011 USDA Animal and Plant Health Inspection Service)
- State and federal statutes
- Annual, inspectors general and Audit reports
- Check out libraries and archives
- Think creatively – beyond paper documents (calendars, phone call logs, expenses, budgets, etc.)

# Federal information

- ① [Fedstats](#) – A to Z list of statistical data from the government.
- ① [Data.gov](#) – Searchable catalogs of government data.
- ① [USAspending.gov](#) – Federal spending data including grants, loans, contracts, etc.
- ① [FedSpending.org](#) – Includes recovery.



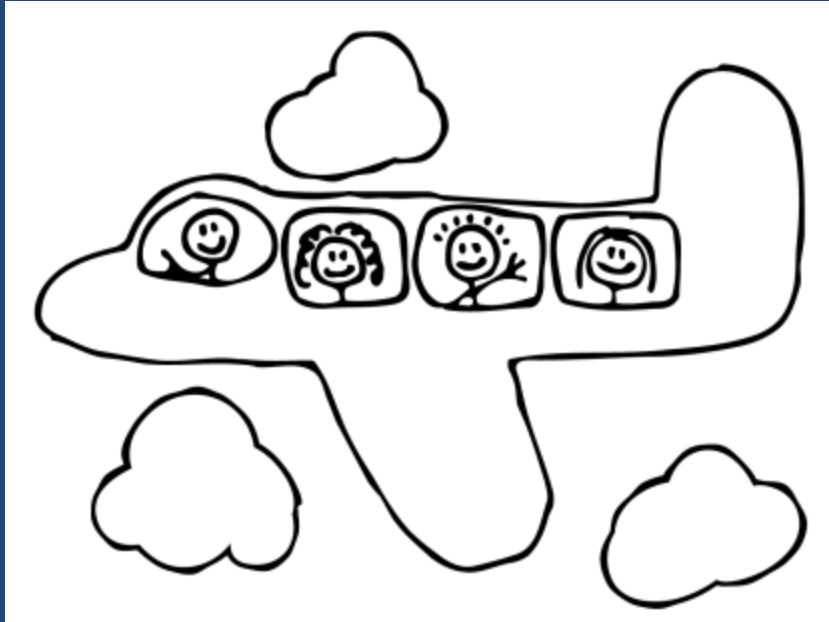
## Searchable and downloadable data

- ⦿ Department of Labor Enforcement Data
- ⦿ OSHA workplace safety inspections
- ⦿ Federal contracts (FPDS) (ex. Border protection)
- ⦿ Recalls (look at tabs for topics)
- ⦿ Economy at a Glance from the Bureau of Labor Statistics
- ⦿ Local Area Unemployment Statistics from the BLS
- ⦿ EPA environmental data searches

- ◎ Fatality Analysis Reporting System data (FARS)
- ◎ National sex offender registry (Department of Justice)
- ◎ Transportation data from BTS
- ◎ Aircraft data (service difficulty reports, on-time data, aircraft registry, etc.)
  - NTSB Aviation Accidents (and more)
  - FAA accidents/incidents and service difficulty reports (SDRs)
  - Flight tracker from flightaware
  - On-time statistics

- ⦿ Nursing home, hospital and home health data from Medicare
- ⦿ Firearms data from the ATF
- ⦿ Migration data from the IRS
- ⦿ Tax data from IRS (tax exempt orgs)
- ⦿ Mine Safety and Health Administration (MSHA)
- ⦿ National Agricultural Statistics Service

# Aviation for consumers



## Consumer Reports

(pdf report)

- On-time by airport (pg 7)
- Departure by time (pg 13)
- Chronically delayed flights (pg 15)
- Tarmac delays (pg 25)
- Mishandled baggage (pg 29)
- Injuries and deaths involving pets (49)

## From Bureau of Transportation statistics

(Downloadable to Excel)

- Flight delays
- Bag fees by airline
- Chronically delayed flights
- Cancelled flights

# The Census

- ◉ [American FactFinder2](#) and [Quickfacts](#) for profiles of states, cities and towns.
- ◉ [1940 Census](#) – Search by location then look for names.
- ◉ Census 2010 and 2000 comparisons from [IRE](#)
- ◉ Census resources from [IRE](#) and [ASU](#)

**CENSUS.IRE.ORG**

Tables from SF1 (What's SF1?)

P1. TOTAL POPULATION  
 P2. URBAN AND RURAL  
 P3. RACE  
 P4. HISPANIC OR LATINO ORIGIN  
 P5. HISPANIC OR LATINO ORIGIN BY RACE  
 P6. RACE (TOTAL RACES TALLIED)  
 P7. HISPANIC OR LATINO ORIGIN BY RACE (TOTAL RACES TALLIED)  
 P8. RACE  
 P9. HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE  
 P10. RACE FOR THE POPULATION 18 YEARS AND OVER  
 P11. HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE FOR THE POPULATION 18 YEARS AND OVER  
 P12. SEX BY AGE  
 P12A. SEX BY AGE (WHITE ALONE)  
 P12B. SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)  
 P12C. SEX BY AGE (AMERICAN INDIAN AND ALASKA NATIVE ALONE)

Make table browser

Add another geography to compare or download as CSV PDF

### P1. TOTAL POPULATION

Universe: Total population

	2000	2030	Delta	% change	2000	2030	Delta	% change
Total	700,820	795,225	94,405.00	13.47	1,737,034	1,931,240	194,215.00	11.15

### P2. URBAN AND RURAL

Universe: Total population

	2000	2030	Delta	% change	2000	2030	Delta	% change
Total	700,820	795,225	94,405.00	13.47	1,737,034	1,931,240	194,215.00	11.15
Urban:	0	0			0	0		
Inside urbanized areas	0	0			0	0		
Inside urban clusters	0	0			0	0		
Rural	0	0			0	0		
Not defined for this file	700,820	795,225	94,405.00	13.47	1,737,034	1,931,240	194,215.00	11.15

# State and local data

- ① State Government Databases wiki from the American Library Association
- ① BRB Publications links to public records sites
- ① Portico

# Tips that data exist

- ⦿ Major Information Systems – request a list from government agency: example US Marshall’s list [here](#).
- ⦿ Something in a “chart” format may mean that a database exists: example California West Nile Virus [here](#).

# More tips...

- ⦿ Online forms to submit information: example Wisconsin Insurance Complaint form [here](#).
- ⦿ Actually read a web site – don't let it force you into predefined roles: example North Carolina Department of Revenue [here](#). (site map, search)
  - Search for keywords related to data: download, database, information system, submit, searchable, inspection, enforcement, Excel, etc.



# Advanced searches

## ◎ Search by domain:

- .gov – government sites (other types: .edu, .org, .com, .net, etc.)
- Example – search for “Gulf oil spill” 2010 and limit to site:.gov

## ◎ Search by file type:

- .xls for spreadsheets; .txt, .csv for text files; .mdb, .dbf for databas files
- Example – Same search as above but limit to filetype:xls

## ◎ Search for words within a URL:

- Example: ftp, download

# Fun with url's

- ⦿ Back up the URL from results of a successful search to see what else the site might have.
- ⦿ Play around with ftp

# Grab someone's Twitter history

- Use this URL:  
*[http://twitter.com/statuses/user\\_timeline/account.xml?count=n](http://twitter.com/statuses/user_timeline/account.xml?count=n)*
- Replace “account” with the Twitter handle and the ‘n’ with the number of Tweets you’d like returned:

 [http://twitter.com/statuses/user\\_timeline/charliesheen.xml?count=500](http://twitter.com/statuses/user_timeline/charliesheen.xml?count=500)

- XML data appears. Save it and open it in Excel.

```
- <statuses type="array">
  - <status>
    <created_at>Mon Mar 21 01:14:20 +0000 2011</created_at>
    <id>49639948269273088</id>
    - <text>
      the warlock's carry-on. c http://twitpic.com/4bmurw
    </text>
    - <source>
      <a href="http://stone.com/Twittelator" rel="nofollow">Twittelator</a>
```

	id	text
34	45240371701219300	#Fastball; The Warlock is hungry. Hungry for corporate flesh. Live tonite - 7p PST. "Building the perfect torpedo" my manifesto @ustream ybw
35	44865655987638300	I'm looking to hire a #winning INTERN with #TigerBlood. Apply here - <a href="http://bit.ly/hykQQF">http://bit.ly/hykQQF</a> #TigerBloodIntern #internship #ad
36	44857523857129500	fastball; Cuban, a mad genius, proclaimed every brain cell in the Warlock brain 2 B a weapon of mass destruction. U've been warned; by him.

- Results include Tweets, retweets, id and more
- Update using the last ID in your list and this URL:  
[http://twitter.com/statuses/user\\_timeline/account.xml?since\\_id=ID](http://twitter.com/statuses/user_timeline/account.xml?since_id=ID)

# Downloading data

- In Excel format – example Illinois Sex Offender Registry [here](#).
- From the web – example California West Nile Virus [here](#).
- In text format – example from data.gov search on “inspection” [here](#).
- Web scraping – Google docs, etc.

Feel free to contact me with any questions:  
[jaimi@ire.org](mailto:jaimi@ire.org)