

Looking for Big Data (BD) in the “Wild”

Tidewater Big Data Enthusiasts
Chuck Cartledge
Developer

April 23, 2020 at 12:18 Noon

Contents

List of Tables	ii
List of Figures	ii
1 Introduction	1
2 Ways to get data	2
2.1 Create your own	2
2.2 Download a file	5
2.3 Download using an Application Program Interface (API)	8
3 Selected Big Data Sources	25
3.1 Aggregator	26
3.2 Aviation	41
3.3 Developers	51
3.4 Education	56
3.5 General	58
3.6 Geographic information	69
3.7 Geographic location	73
3.8 Government	74
3.9 Social	105
3.10 Weather	110
3.11 Zip code	123
4 System performance	124

List of Tables

1	Contents of the file Medicare_Provider_Util_Payment_PUF_CY2013.zip. . . .	8
---	---	---

List of Figures

1	Nicholas Felton’s annual report.	3
2	Aaron Parecki records his GPS position every 5 seconds.	4
3	Centers for Medicare and Medicaid Services download site.	6
4	Medicare data file viewed normally.	7
5	Medicare data file shown with tabs.	7
6	Project Gutenberg home page.	9
7	Project Gutenberg’s copy of Romeo and Juliet.	10
8	Flickr’s API page.	12
9	Flickr’s flickr.photos.search API link.	13
10	Flickr’s flickr.photos.getWithGeoData API page.	14
11	Flickr’s API key page.	15
12	Flickr’s API key type selection page.	16
13	Flickr’s application description page.	17
14	Flickr’s API exploration link.	18
15	Flickr’s API exploration page.	20
16	Flickr’s API exploration page with bounding box data.	21
17	Flickr’s API exploration calling method.	22
18	Flickr’s API exploration with data and XML format.	23
19	Flickr’s API exploration with data and JSON format.	24
20	70+ websites to get large data repositories for free home page.	26
21	Anywhere that I could download a sample database? : learnprogramming home page.	27
22	Data Portals home page.	28
23	Datasets Archive home page.	29
24	Datasets for Data Mining and Data Science home page.	30
25	enigma home page.	31
26	Finding Data on the Internet home page.	32
27	Linked Data - Connect Distributed Data across the Web home page.	33
28	Oxford Internet Institute University of Oxford home page.	34
29	Real Time Data Resources home page.	35
30	Thirty Thousand Feet home page.	36

31	Top 10 Weather APIs home page.	37
32	University Corporation for Atmospheric Research home page.	38
33	Where can I find large datasets open to the public? - Quora home page.	39
34	Wikipedia:Database download home page.	40
35	Flight Aware home page.	41
36	Flight Stats home page.	42
37	Flight View home page.	43
38	OAG (Official Airline Guide) home page.	44
39	QPX Express API home page.	45
40	Travel Boutique home page.	47
41	Travel Fusion home page.	49
42	Wego Flights home page.	50
43	GeoCommons home page.	51
44	Global Data. Local Context home page.	52
45	OpenStreetMap home page.	53
46	Quandl Financial and Economic Data home page.	54
47	TIGER Products home page.	55
48	Delve Datasets home page.	56
49	Statistics Online Computational Resource (SOCR) home page.	57
50	Bing home page.	58
51	Blog - AriLamstein.com home page.	59
52	Dogpile home page.	60
53	DuckDuckGo home page.	62
54	Google home page.	63
55	Our World in Data home page.	64
56	Stanford Large Network Dataset Collection home page.	65
57	UCI Machine Learning Repository home page.	66
58	WolframAlpha home page.	67
59	Yahoo home page.	68
60	GeoNames home page.	69
61	Open Source Data and APIs — MaxMind home page.	70
62	Places API - CityGrid V2 - CityGrid Media Documentation home page.	72
63	MaxMind Open Source Data and APIs home page.	73
64	Austin Texas Data Portal home page.	74
65	Bureau of Economic Analysis home page.	75
66	Bureau of Labor Statistics home page.	76
67	Census Bureau Economic Statistics home page.	77
68	City Maps - Robert Wood Johnson Foundation home page.	78
69	Datasets - data.sa.gov.au home page.	79
70	Earthquake Hazards Program home page.	80

71	FAA Web Services, Airport Service home page.	81
72	Federal Reserve Bank of St. Louis, Economic Research home page.	82
73	Gapminder Data home page.	83
74	General Social Survey home page.	85
75	Massachusetts Bay Transportation Authority home page.	86
76	National Aeronautics and Space Administration, Goddard Institute for Space Studies home page.	87
77	National Flight Data Center (NFDC) home page.	88
78	National Longitudinal Surveys — A Program of the U.S. Bureau of Labor Statistics home page.	89
79	NYC Open Data home page.	90
80	Parking in San Francisco, CA home page.	91
81	SNAP Retailer Locator home page.	92
82	Social Security Administration home page.	93
83	The National Archives home page.	94
84	THE National Map home page.	95
85	The National UFO Reporting Center home page.	96
86	The World Bank home page.	97
87	UK Government Open Data Portal home page.	98
88	United States Department of Agriculture, Economic Research Service home page.	99
89	USDA Food Composition Database home page.	100
90	US Department of Transportation home page.	101
91	US Government Open Data Portal home page.	102
92	USGovXML.com: US Government Web Services and XML Data Sources. home page.	103
93	Washington Metropolitan Area Transit Authority API home page.	104
94	American Customer Satisfaction Index home page.	105
95	American Customer Satisfaction Index home page.	106
96	Linkedin API home page.	107
97	Meetup API home page.	108
98	YELP API home page.	109
99	AccuWeather home page.	110
100	Aeris Weather home page.	111
101	Dark Sky Forecast home page.	112
102	National Data Buoy Center home page.	113
103	National Weather Service home page.	114
104	OpenWeatherMap home page.	115
105	Real-time weather forecasts home page.	116
106	WeatherBug home page.	117

107	Weather Source home page.	118
108	Weather Underground home page.	119
109	World Weather Online home page.	120
110	World Weather OnLine home page.	121
111	Zip Codes home page.	123
112	How long it took to create the home page images.	124

1 Introduction

One way to view things in the Big Data (BD) world is as a collaboration by three individuals or groups [2]. These are:

- *Data visionary*: the one with the vision of how different data could be combined,
- *Data scientist*: the one with the technical and analytical skills to unify disparate sources of data, and
- *Data broker*: the one with raw data

A simple example will clarify the things that each of these groups brings to the table.

- *Statement by data visionary*: What is the relationship between: (1) US annual change in inflation, (2) US annual adjusted unemployment rate, (3) political party in US House of Representatives and Congress, and (4) US major wars and conflicts for the 20th and 21st centuries? The goal to be able to make a statement like: When party X is in control and we are/are not involved in a major conflict, then we can expect unemployment to do something and inflation to do something.
- *Data broker*:
 1. Inflation data is available from the US Bureau of Labor Statistics¹
 2. Unemployment level data is available from the US Bureau of Labor Statistics²
 3. A “good enough” listing of how many members of each part for each house of congress³
 4. A “good enough” listing of US major wars and conflicts⁴
- *Data scientist*: how to display all the data on a single X-Y plot, where time is on the X axis? The intent being to be able to look up from the X axis and see if the US was in a conflict, which party was controlling the House and the Senate, and what the change in unemployment and inflation was from the previous year.

In the following sections, we’ll talk about becoming your own *data broker*.

¹<http://data.bls.gov/pdq/SurveyOutputServlet>

²<http://data.bls.gov/cgi-bin/surveymost?r4>

³http://en.wikipedia.org/wiki/Party_divisions_of_United_States_Congresses

⁴http://en.wikipedia.org/wiki/List_of_wars_involving_the_United_States

2 Ways to get data

2.1 Create your own

Perhaps the most overlooked source of Big Data is you.

Everyday you create data, some of which is captured, most is lost. The data that is captured can tell a lot about you, and can be displayed in interesting ways. Nicholas Felton has been collecting and publishing personal data since 2005 [1] (see Figure 1). Your smartphone can provide all sorts of data about where you are when and how fast you're going. Aaron Parecki is the co-founder of IndieWebCamp, and maintains oauth.net. He is known for having tracked his location at 5 second intervals since 2008 [3] (see Figure 2). If you are wearing a Fitbit (or other wearable sensor), then you are creating data all the time. Now it is a matter of collecting and accessing it. Locating, understanding, and manipulating the data can be a challenge.

Even if you aren't wearing a wearable device, most likely you are carrying one. Your smartphone is carrying an array of sensors, including:

1. Motion sensors that can tell the difference between waling and driving,
2. A barometer for measuring atmospheric pressure,
3. A gesture sensor that detects hand movements through infrared rays,
4. Gyroscope to measure acceleration,
5. Magnetometer to measure magnetic lines of flux,
6. GPS to tell where you are around the world,
7. WIFI to connect to the world, and also to tell how close you are to a broadcast station or router,
8. Camera(s) to see,
9. Microphone(s) to listen,
10. Speaker(s) to speak,
11. Temperature and pressure (on the screen).

Undoubtedly, there will be more as technology gets smaller, as they become more energy efficient, and people begin to want more and more. One immediate advantage (depending on where you live), will be that you can have your insurance rate reflect how you really drive.



Figure 1: Nicholas Felton's annual report.

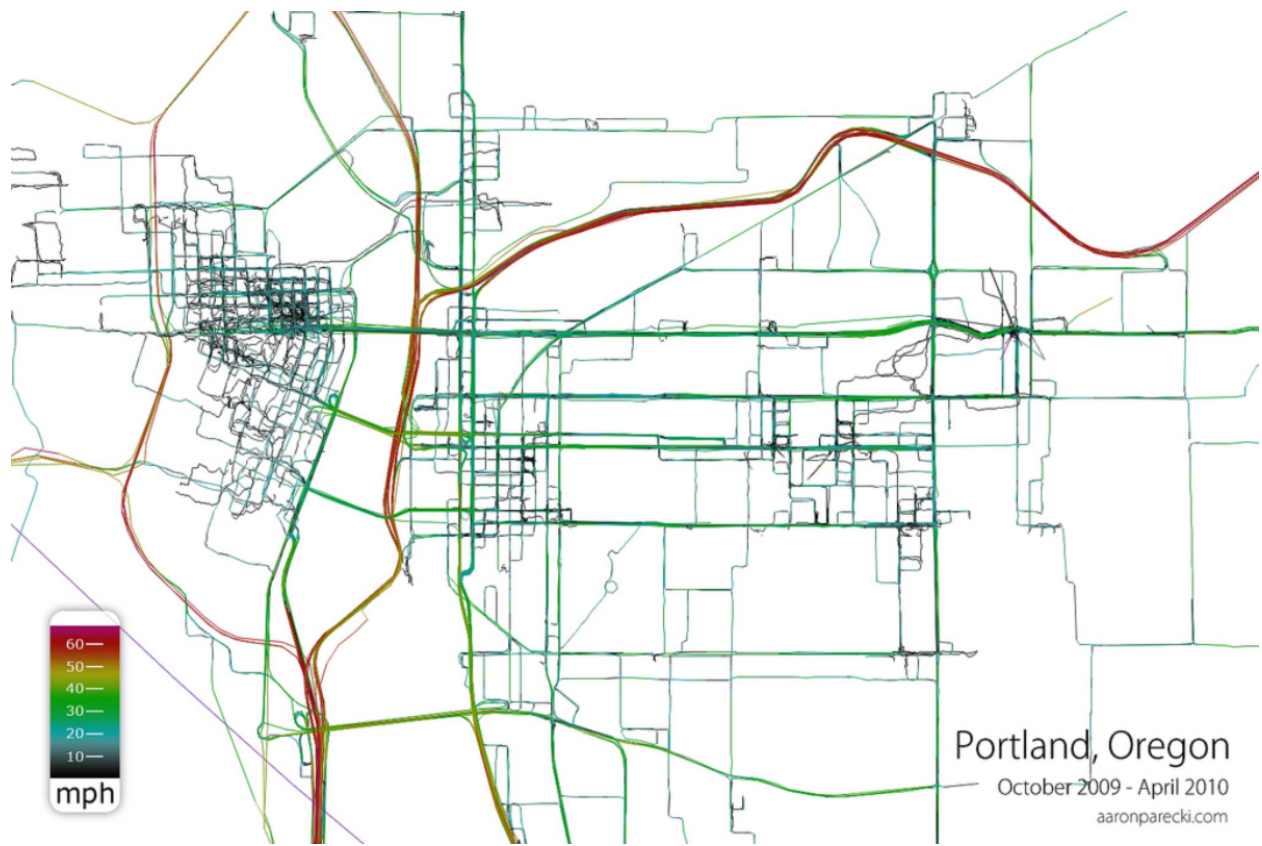


Figure 2: Aaron Parecki records his GPS position every 5 seconds.

An app running on your smartphone could communicate with the insurance company in real-time to report on your current driving activities⁵. Currently this is an “opt-in” application, but what could happen when the “opt-inners” outnumber the “opt-outters”? What will that say about your own perception of your driving habits? It could almost be Orwellian.

2.2 Download a file

Another way to acquire enough BD to have “fun” with is to download a file from somewhere that contains things of interest. Sometimes it is restively easy to find somewhere to download the file, other times it may take a while. Once you have located the file, you’ll have to process it in some manner. The exact processing details will depend on how the file is laid out, and what tools you have available. All of this sounds very esoteric, and unfulfilling so we’ll work through a couple of examples. The first will be to download a file from Medicare with tab separated values (TSV)⁶. The second will be a unformatted text file.

The Centers for Medicare and Medicaid Services⁷ makes available a vast array of data relating to all their programs. As a specific example of a TSV, we will look at Medicare Payments during calendar year 2013 (see Figure 3)⁸. We are interested in the Medicare Physician and Other Supplier PUF, CY2013, tab delimited format file for download. Some particulars about the file Medicare_Provider_Util_Payment_PUF_CY2013.zip.

1. It is 497,014,400 bytes of compressed data.
2. It contains three other files (see Table 1).
3. The data file has a header record.
4. The data fields are delimited by tab characters that are normally invisible (see Figure 4).
5. The a single tab is between each data field. The editor can make the tabs visible (see Figure 5).

In the final analysis: the file is large, it is well structured, most modern languages will be able to process the file without trouble.

⁵<https://www.washingtonpost.com/news/the-switch/wp/2016/01/04/the-big-data-of-bad-driving-and-how-insurers-plan-to-track-your-every-turn/>

⁶TSV files are a subset of a general class of files called “delimited text” files. All delimited files strive to identify and use a particular character (either a TAB, or a SPACE, or a COMMA, or DOUBLE QUOTES) that the originator believes will not occur normally in the data.

⁷<https://www.cms.gov/>

⁸<https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Physician-and-Other-Supplier2013.html>



Figure 3: Centers for Medicare and Medicaid Services download site.

```

chuck@drone: ~/Downloads: arch Terminal Help
File Edit Options Buffers Tools Help
NPI      NPPES_PROVIDER_LAST_ORG_NAME      NPPES_PROVIDER_FIRST_NAME      NPPES_PROVIDER_MI      NPPES_CREDENTIALS      $
000000001      CPT copyright 2012 American Medical Association. All Rights Reserved.      $
1003000126      ENKESHAFI      ARDALAN      M.D.      M      I      900 SETON DR      CUMBERLAND      215021$
1003000126      ENKESHAFI      ARDALAN      M.D.      M      I      900 SETON DR      CUMBERLAND      215021$
1003000126      ENKESHAFI      ARDALAN      M.D.      M      I      900 SETON DR      CUMBERLAND      215021$
1003000126      ENKESHAFI      ARDALAN      M.D.      M      I      900 SETON DR      CUMBERLAND      215021$
1003000126      ENKESHAFI      ARDALAN      M.D.      M      I      900 SETON DR      CUMBERLAND      215021$
1003000126      ENKESHAFI      ARDALAN      M.D.      M      I      900 SETON DR      CUMBERLAND      215021$
1003000126      ENKESHAFI      ARDALAN      M.D.      M      I      900 SETON DR      CUMBERLAND      215021$
1003000134      CIBULL THOMAS      L      M.D.      M      I      2650 RIDGE AVE      EVANSTON HOSPITAL      EVANSTON      $
1003000134      CIBULL THOMAS      L      M.D.      M      I      2650 RIDGE AVE      EVANSTON HOSPITAL      EVANSTON      $
1003000134      CIBULL THOMAS      L      M.D.      M      I      2650 RIDGE AVE      EVANSTON HOSPITAL      EVANSTON      $
1003000134      CIBULL THOMAS      L      M.D.      M      I      2650 RIDGE AVE      EVANSTON HOSPITAL      EVANSTON      $
1003000134      CIBULL THOMAS      L      M.D.      M      I      2650 RIDGE AVE      EVANSTON HOSPITAL      EVANSTON      $
1003000134      CIBULL THOMAS      L      M.D.      M      I      2650 RIDGE AVE      EVANSTON HOSPITAL      EVANSTON      $
1003000134      CIBULL THOMAS      L      M.D.      M      I      2650 RIDGE AVE      EVANSTON HOSPITAL      EVANSTON      $
1003000142      KHALIL RASHID      M.D.      M      I      4126 N HOLLAND SYLVANIA RD      SUITE 220      TOLEDO$
1003000142      KHALIL RASHID      M.D.      M      I      4126 N HOLLAND SYLVANIA RD      SUITE 220      TOLEDO$
1003000142      KHALIL RASHID      M.D.      M      I      4126 N HOLLAND SYLVANIA RD      SUITE 220      TOLEDO$
1003000142      KHALIL RASHID      M.D.      M      I      4126 N HOLLAND SYLVANIA RD      SUITE 220      TOLEDO$
1003000142      KHALIL RASHID      M.D.      M      I      4126 N HOLLAND SYLVANIA RD      SUITE 220      TOLEDO$
1003000142      KHALIL RASHID      M.D.      M      I      4126 N HOLLAND SYLVANIA RD      SUITE 220      TOLEDO$
1003000142      KHALIL RASHID      M.D.      M      I      4126 N HOLLAND SYLVANIA RD      SUITE 220      TOLEDO$
-UUU:---F1 temp.txt      Top L??      (Text Archive pair) -----
No further undo information

```

Figure 4: Medicare data file viewed normally.

```

chuck@drone: ~/Downloads: arch Terminal Help
File Edit Options Buffers Tools Help
NPI      NPPES_PROVIDER_LAST_ORG_NAME      NPPES_PROVIDER_FIRST_NAME      NPPES_PROVIDER_MI      NPPES_CREDENTIALS      $
000000001      CPT copyright 2012 American Medical Association. All Rights Reserved.      $
1003000126      ENKESHAFI      ARDALAN      M.D.      M      I      900 SETON DR      CUMBERLAND      215021$
1003000126      ENKESHAFI      ARDALAN      M.D.      M      I      900 SETON DR      CUMBERLAND      215021$
1003000126      ENKESHAFI      ARDALAN      M.D.      M      I      900 SETON DR      CUMBERLAND      215021$
1003000126      ENKESHAFI      ARDALAN      M.D.      M      I      900 SETON DR      CUMBERLAND      215021$
1003000126      ENKESHAFI      ARDALAN      M.D.      M      I      900 SETON DR      CUMBERLAND      215021$
1003000126      ENKESHAFI      ARDALAN      M.D.      M      I      900 SETON DR      CUMBERLAND      215021$
1003000126      ENKESHAFI      ARDALAN      M.D.      M      I      900 SETON DR      CUMBERLAND      215021$
1003000134      CIBULL THOMAS      L      M.D.      M      I      2650 RIDGE AVE      EVANSTON HOSPITAL      EVANSTON      $
1003000134      CIBULL THOMAS      L      M.D.      M      I      2650 RIDGE AVE      EVANSTON HOSPITAL      EVANSTON      $
1003000134      CIBULL THOMAS      L      M.D.      M      I      2650 RIDGE AVE      EVANSTON HOSPITAL      EVANSTON      $
1003000134      CIBULL THOMAS      L      M.D.      M      I      2650 RIDGE AVE      EVANSTON HOSPITAL      EVANSTON      $
1003000134      CIBULL THOMAS      L      M.D.      M      I      2650 RIDGE AVE      EVANSTON HOSPITAL      EVANSTON      $
1003000134      CIBULL THOMAS      L      M.D.      M      I      2650 RIDGE AVE      EVANSTON HOSPITAL      EVANSTON      $
1003000134      CIBULL THOMAS      L      M.D.      M      I      2650 RIDGE AVE      EVANSTON HOSPITAL      EVANSTON      $
1003000142      KHALIL RASHID      M.D.      M      I      4126 N HOLLAND SYLVANIA RD      SUITE 220      TOLEDO$
1003000142      KHALIL RASHID      M.D.      M      I      4126 N HOLLAND SYLVANIA RD      SUITE 220      TOLEDO$
1003000142      KHALIL RASHID      M.D.      M      I      4126 N HOLLAND SYLVANIA RD      SUITE 220      TOLEDO$
1003000142      KHALIL RASHID      M.D.      M      I      4126 N HOLLAND SYLVANIA RD      SUITE 220      TOLEDO$
1003000142      KHALIL RASHID      M.D.      M      I      4126 N HOLLAND SYLVANIA RD      SUITE 220      TOLEDO$
1003000142      KHALIL RASHID      M.D.      M      I      4126 N HOLLAND SYLVANIA RD      SUITE 220      TOLEDO$
1003000142      KHALIL RASHID      M.D.      M      I      4126 N HOLLAND SYLVANIA RD      SUITE 220      TOLEDO$
-UUU:---F1 temp.txt      Top L??      (Text Archive pair) Isearch) -----
I-search: ^I

```

Figure 5: Medicare data file shown with tabs. Some editors go to great lengths to expand each tab character into some number of spaces (often 8) to try and make the text “pretty” on the display.

Length	Date	Time	File name
24,011	27-May-2015	14:47:38	CMS_AMA_CPT_license_agreement.pdf
3,650	14-Apr-2015	13:42:36	Medicare-Physician-and-Other-Supplier-PUF-SAS-Infile.sas
2,209,344,403	9-Apr-2015	16:14:28	Medicare_Provider_Util_Payment_PUF_CY2013.txt

Table 1: Contents of the file Medicare.Provider.Util.Payment.PUF.CY2013.zip. The data is almost 18 months old, and is the most recent available. The payment file is over 2.2Gigabytes in size and has 9,287,878 lines of data.

Project Gutenberg (PG)⁹ (see Figure 6) is a volunteer effort to digitize and archive cultural works, to “encourage the creation and distribution of eBooks.”¹⁰ It was founded in 1971 by Michael S. Hart and is the oldest digital library. Most of the items in its collection are the full texts of public domain books. The project tries to make these as free as possible, in long-lasting, open formats that can be used on almost any computer. As of 3 October 2015, Project Gutenberg reached 50,000 items in its collection. The PG serves as a good source of text for many samples of free text. In this example we will look at the play: Romeo and Juliet (see Figure 7).

Some particulars about the PG version of Romeo and Juliet:

1. It has 5,557 lines.
2. It has 27,424 words.
3. It has 153,666 characters.
4. It has a PG specific header that is 289 lines long.

In the final analysis: the file is small, it is free form text, most modern languages will be able to load the file without trouble. However making sense of the file is a natural language processing (NLP) problem, and some computer languages are better than others when dealing with NLP tasks.

2.3 Download using an Application Program Interface (API)

Conceptually, application program interface (API) is a contract between one piece of software and another. One of these pieces of software will provide a service (commonly called the “server”), to the other (commonly called the “client”). The client provides data using the correct protocols and formats to the server, and the server will “fulfill” the contract by performing the appropriate action. Some actions that a server could perform include:

⁹<https://www.gutenberg.org/>

¹⁰http://www.gutenberg.org/wiki/Gutenberg:Project_Gutenberg_Mission_Statement_by_Michael_Hart



Figure 6: Project Gutenberg home page.

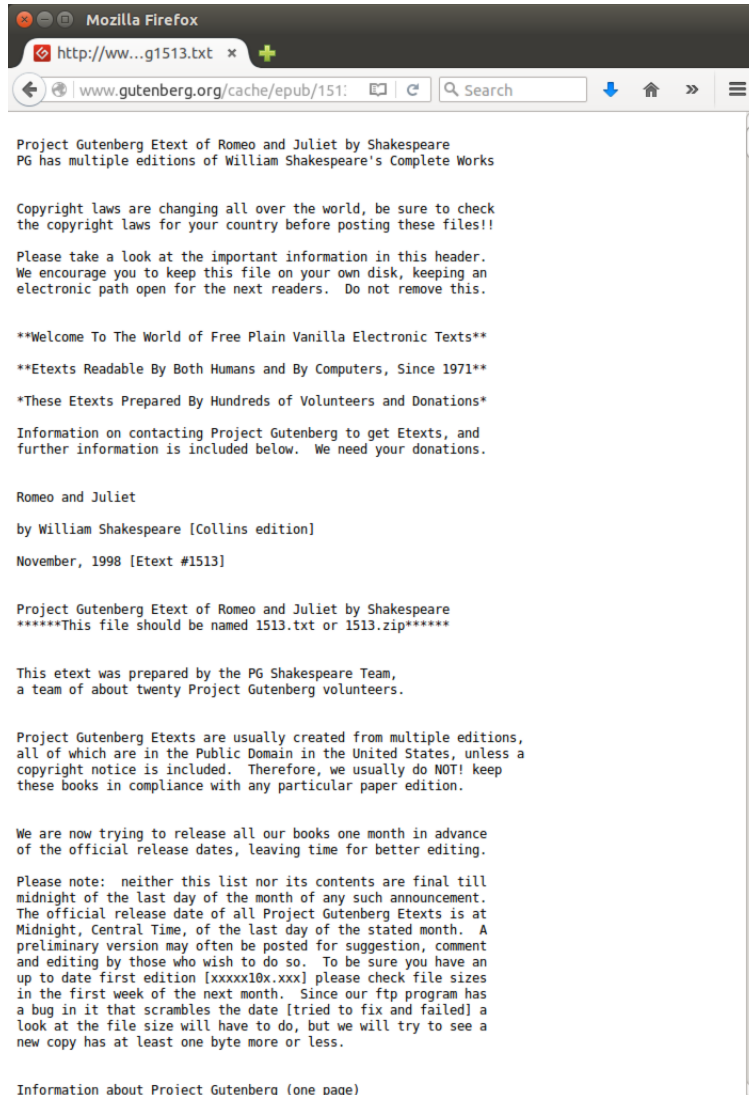


Figure 7: Project Gutenberg's copy of Romeo and Juliet.

- Return a web page when the client (a Web browser, using the hypertext transport protocol (HTTP)), properly requests one, or
- Compute a trigonometric function when the sin button on a calculator is pressed, or
- Encapsulate raw bits from a network interface into data frames as part of the Open System Interconnection (OSI) protocol.

It is the responsibility of the client to understand the details of the contract when requesting something from the server. The only standard that exists across all APIs is a functional one. To wit; if the client meets its side of the contract, then the server must meet its side. The server can unilaterally change the terms of the contract.

As an example, we will look at a moderately complex API to programmatically download pictures from Flickr.

1. Flickr has an extensive API¹¹ to allow applications to interact with Flickr and do all the things that a human would do, plus more (see Figure 8)
2. If you scroll down the Flickr API page, we are interested in using “flickr.photos.search” API (see Figure 9).
3. Flickr’s flickr.photos.search API¹² “contract” has one required value, and wide range of optional values (see Figure 10).
4. The flickr.photos.search API requires a Flickr API key¹³. Applying for the key is completely on-line and is quick and easy.
5. If your interest is commercial, then apply for a commercial key. If your interest is browsing or investigative, then apply for a non-commercial key (see Figure 12).
6. You fill in all the necessary information, and press the SUBMIT key (see Figure 13).
7. Flickr will respond with a page showing the public and private API keys associated with your application (screen not shown). The public key is the one that the flickr.photos.search requires. The secret key is one that you should keep private to yourself and only use when necessary to authenticate yourself to Flickr.
8. The Flickr API page lists all the optional and required arguments with a brief explanation of the type of data associated with the argument. At the bottom of the page is a link to explore how the API operates (see Figure 14).

¹¹<https://www.flickr.com/services/api/>

¹²<https://www.flickr.com/services/api/flickr.photos.search.html>

¹³https://www.flickr.com/services/api/misc.api_keys.html

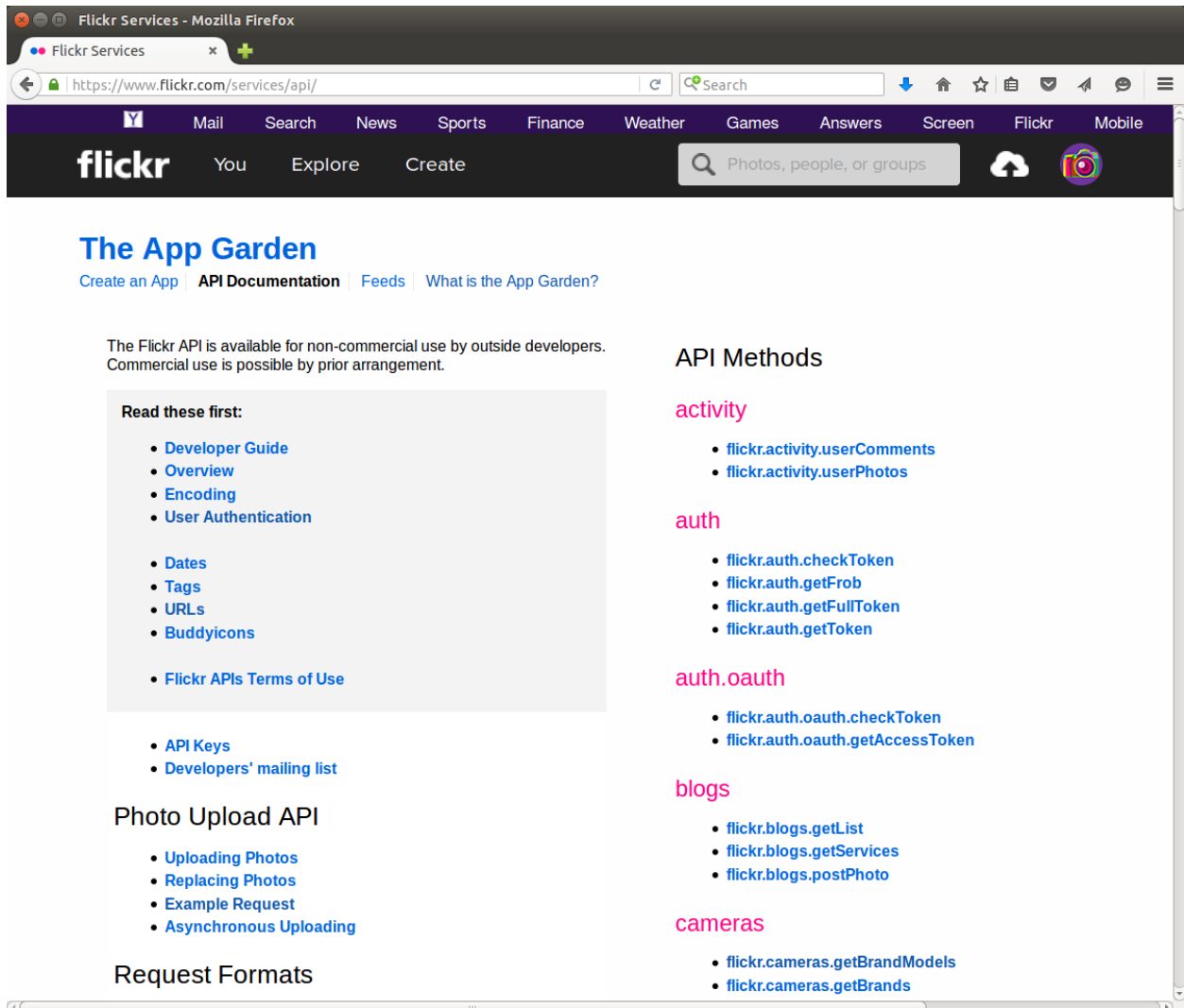


Figure 8: Flickr's API page.

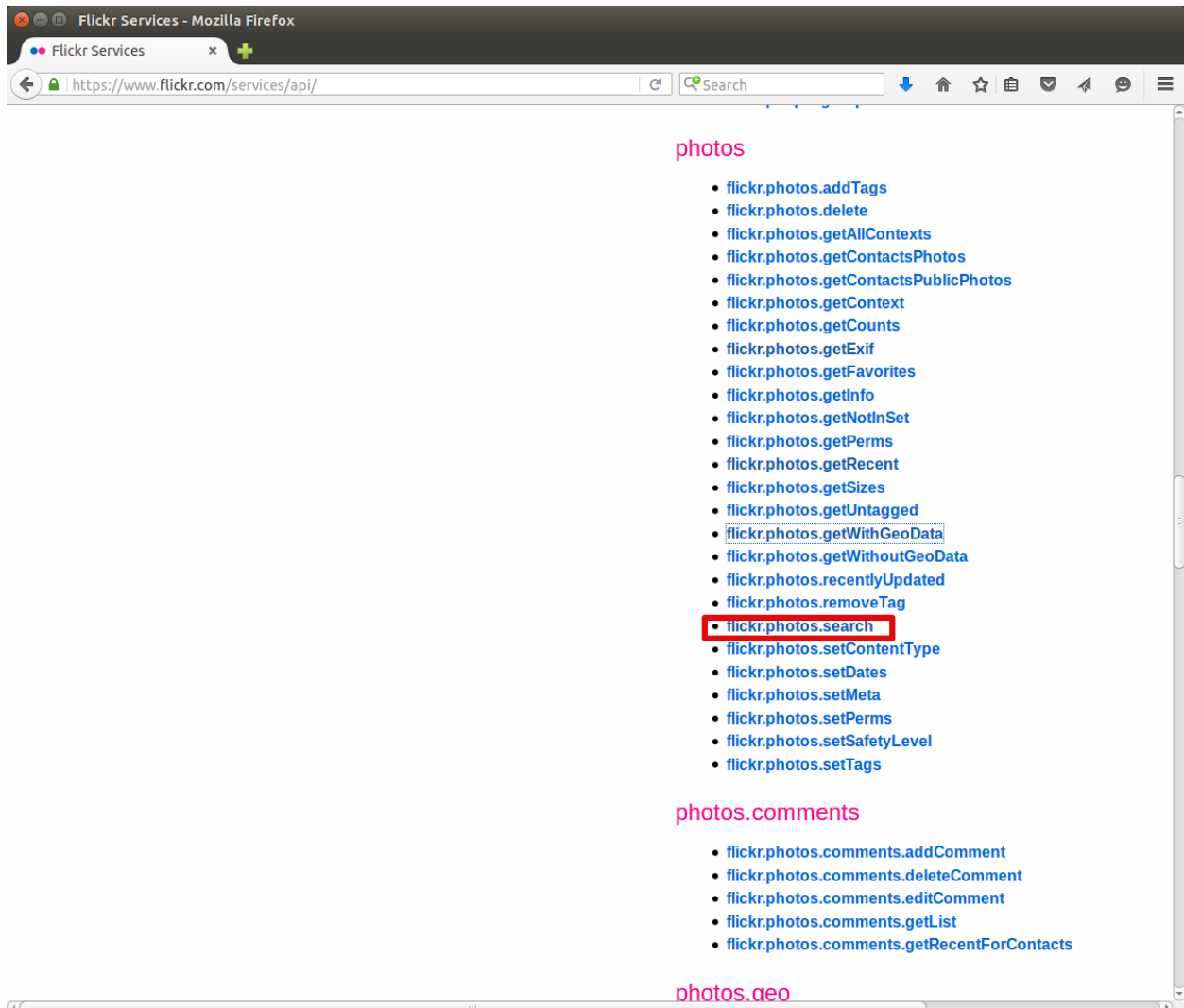


Figure 9: Flickr's flickr.photos.search API link. The link of interest is highlighted in red.

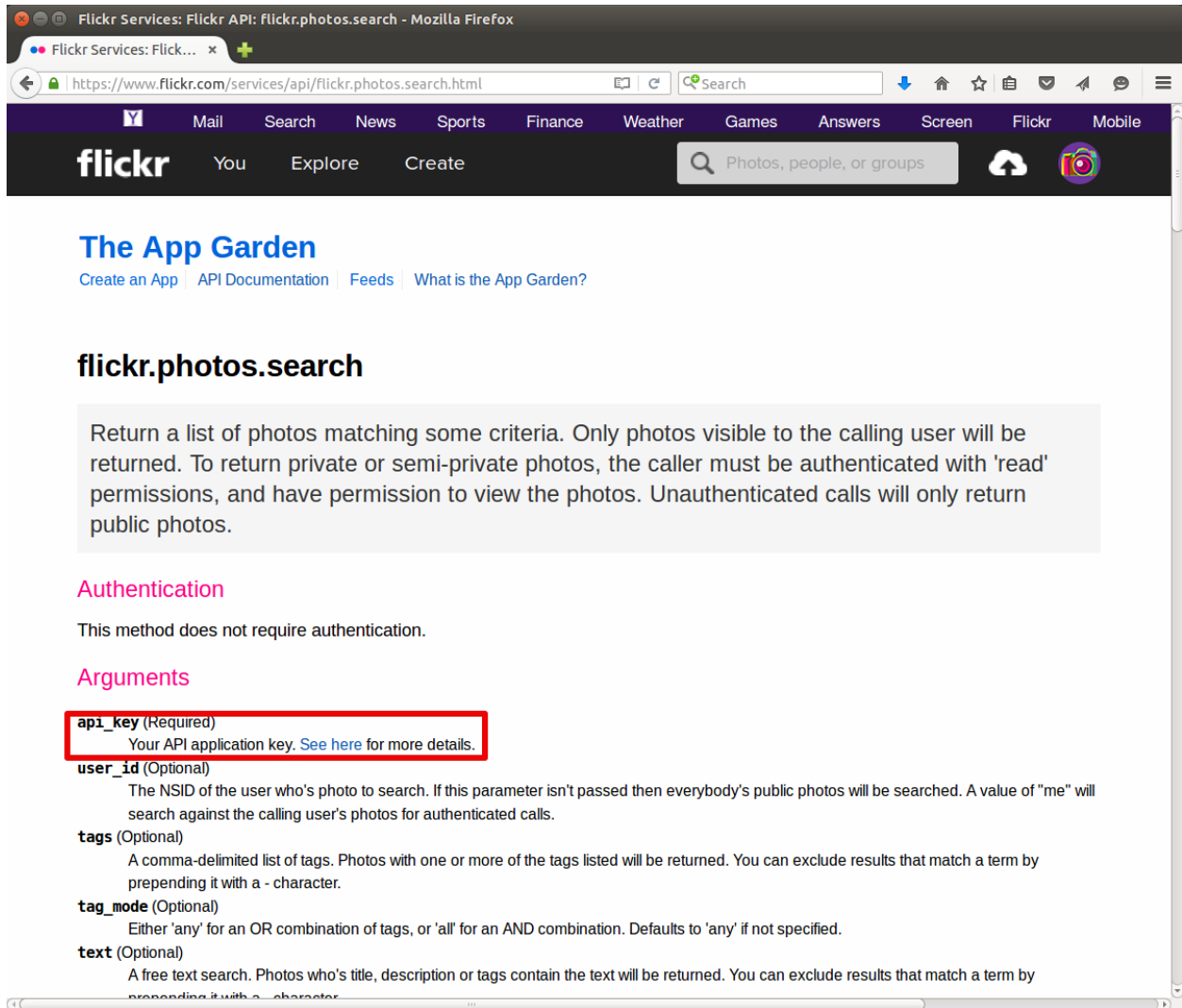


Figure 10: Flickr's `flickr.photos.getWithGeoData` API page. The single required data value is highlighted in red.

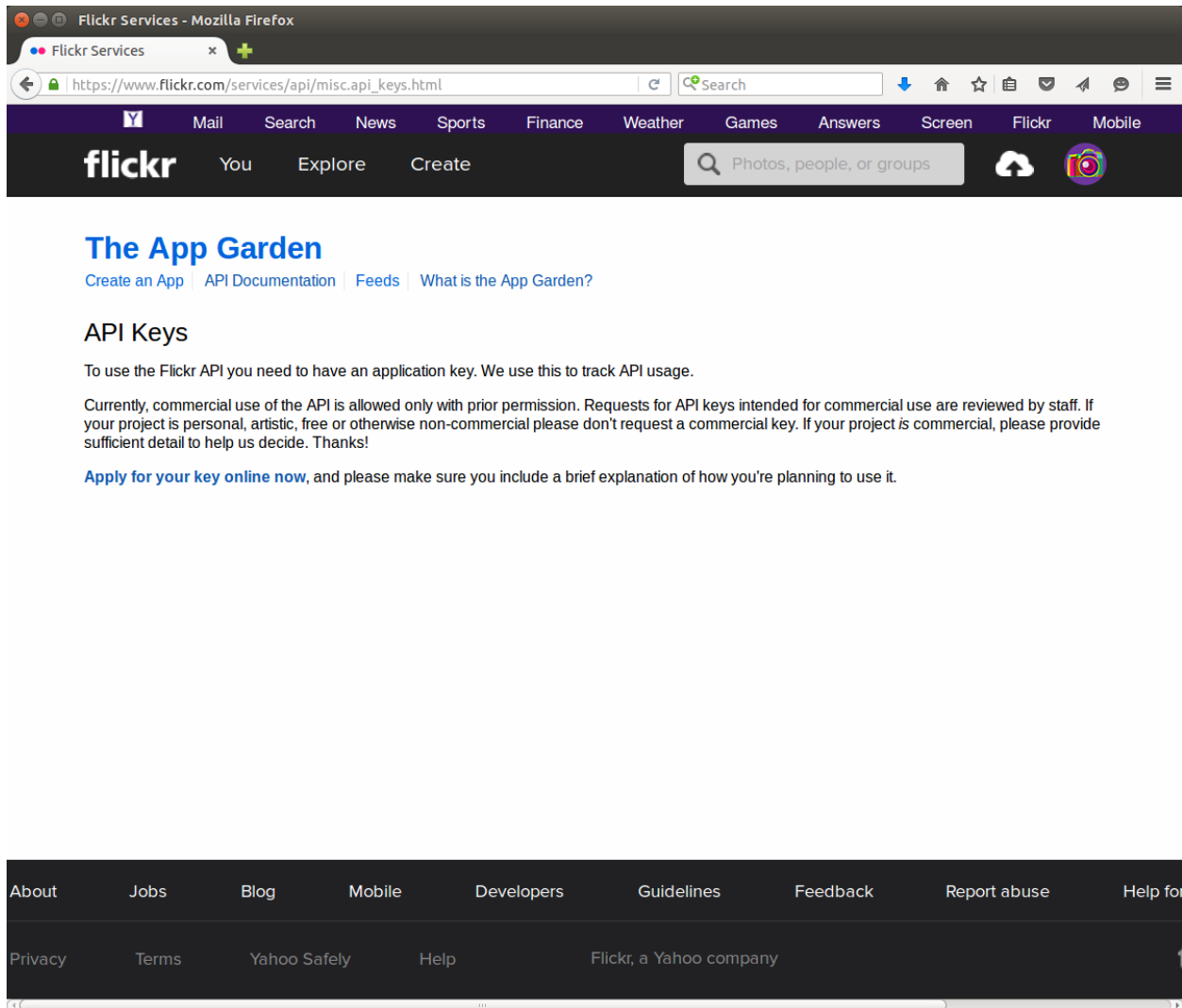


Figure 11: Flickr's API key page.

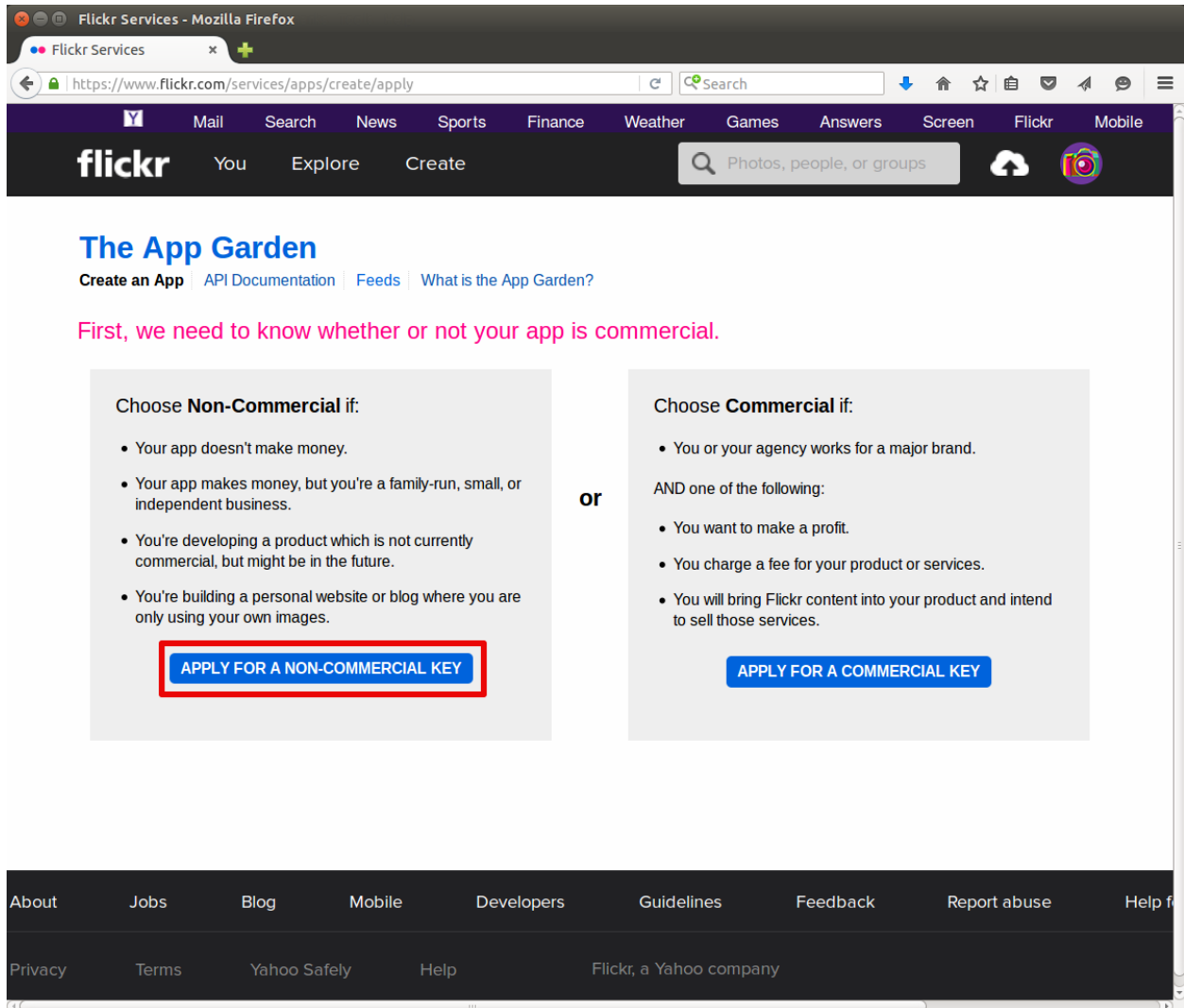


Figure 12: Flickr's API key type selection page. The non-commercial key selection is highlighted.

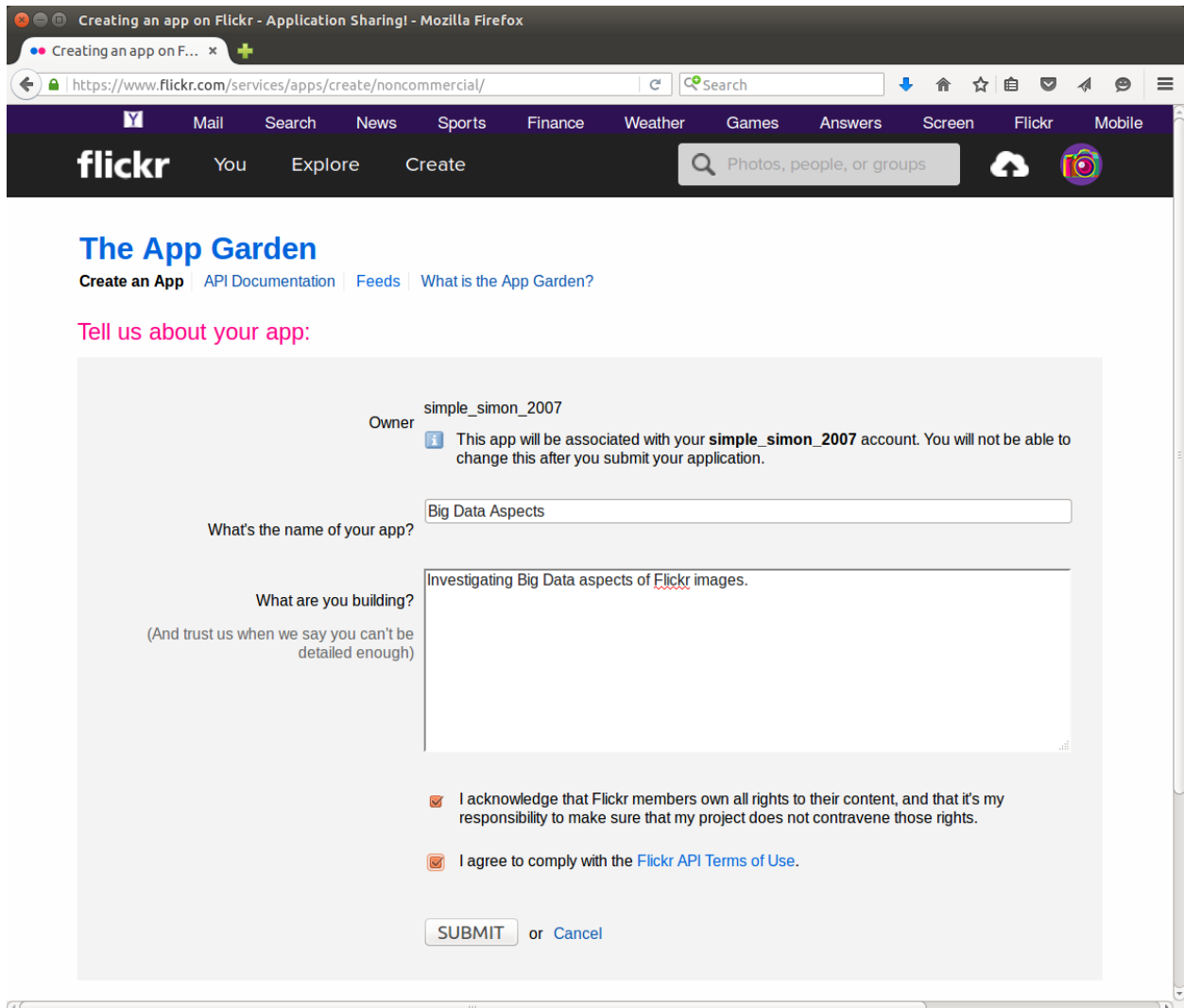


Figure 13: Flickr's application description page.

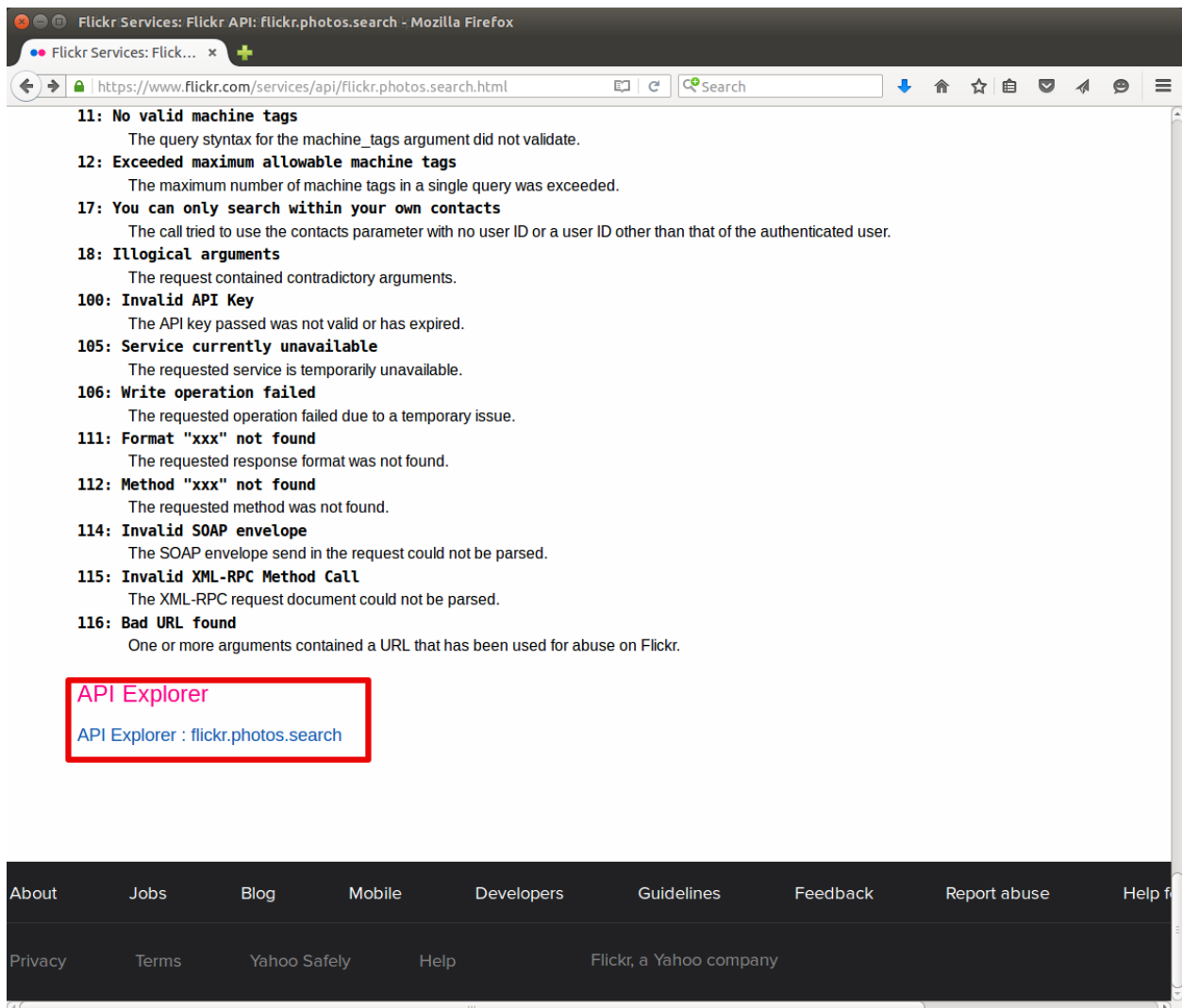


Figure 14: Flickr's API exploration link. The link is highlighted in red.

9. The Flickr API explorer page (see Figure 15)¹⁴ allows you to fill out data for almost all the fields associated with the flickr.photos.search API. The only field that is not available is the API key field. The exploration field uses a default value for the key, so you don't have to enter one.
10. For this exploration, we are looking for pictures tagged with a position between 36 and 37 North, and 75 and 76 West
11. We are now ready to interact with the API and get data back in an XML format (see Figure 17).
12. After pressing the Call Method button, the page is refreshed. At the bottom of the page is the data returned from the call, and the URL used get the data from Flickr (see Figure 18). There are several items of interest in the response window. They include:
 - (a) The number of pages of data that match the request (580 in this example).
 - (b) The data shown comes from a particular page (1 in this example).
 - (c) Each page has a number of picture references (100 in this example).
 - (d) The total number of pictures (57,993 in this example).
13. The same API method call made a few minutes later only selecting JSON (JavaScript Object Notation) as the output format (see Figure 19). The difference between XML (eXtensible Markup Language) is more a matter of how you think about the data and the tools that you have available to manipulate the data rather than the data itself. The values are slightly different than the XML image because time has passed and the picture owners may have removed pictures from Flickr. Remember that Flickr is a semi-real time application because people can change things without notifying you.

Sample flickr JSON and XML response files are attached¹⁵.

¹⁴<https://www.flickr.com/services/api/explore/flickr.photos.search>

¹⁵The JSON file is here. The XML file is here.

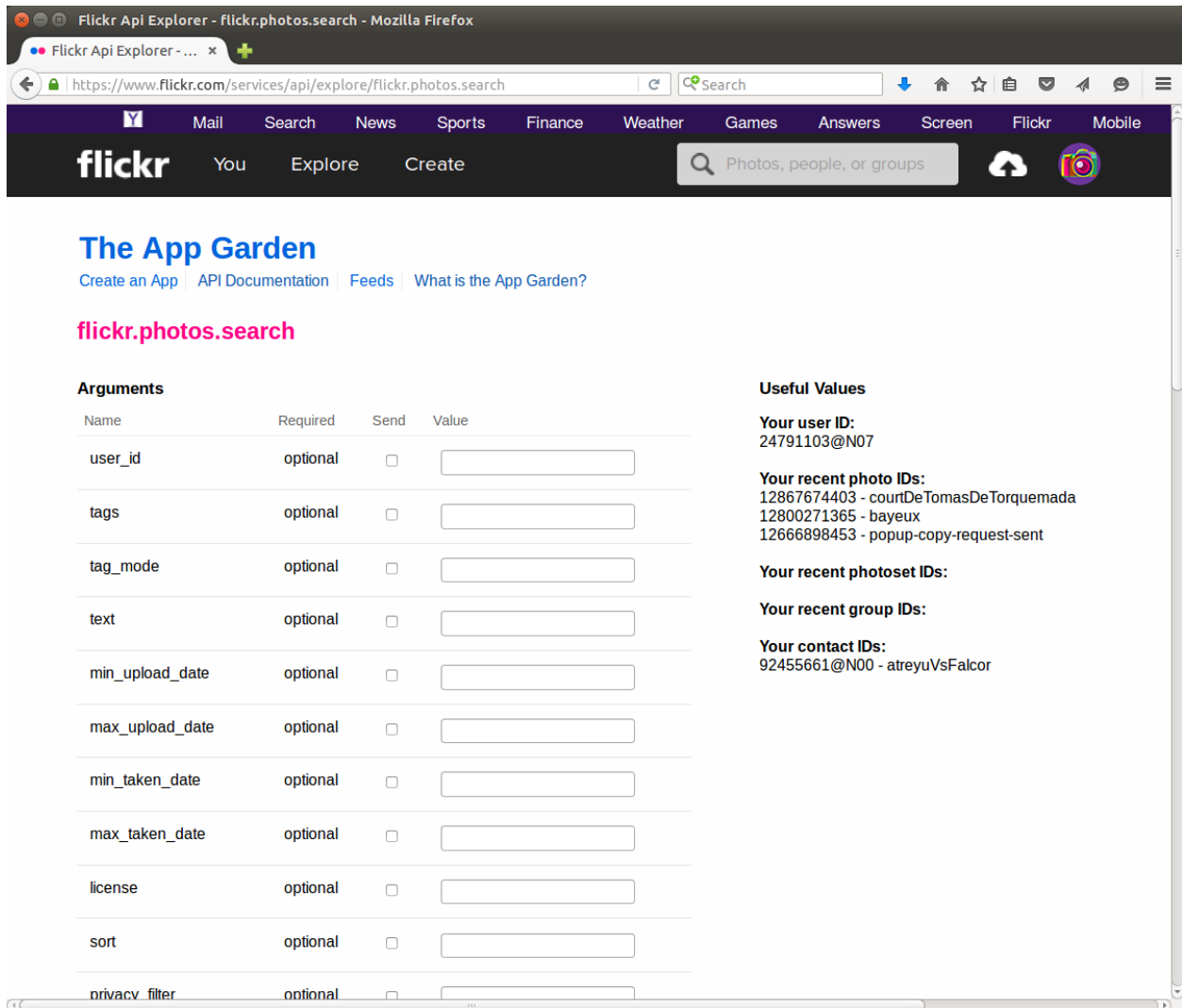


Figure 15: Flickr's API exploration page.

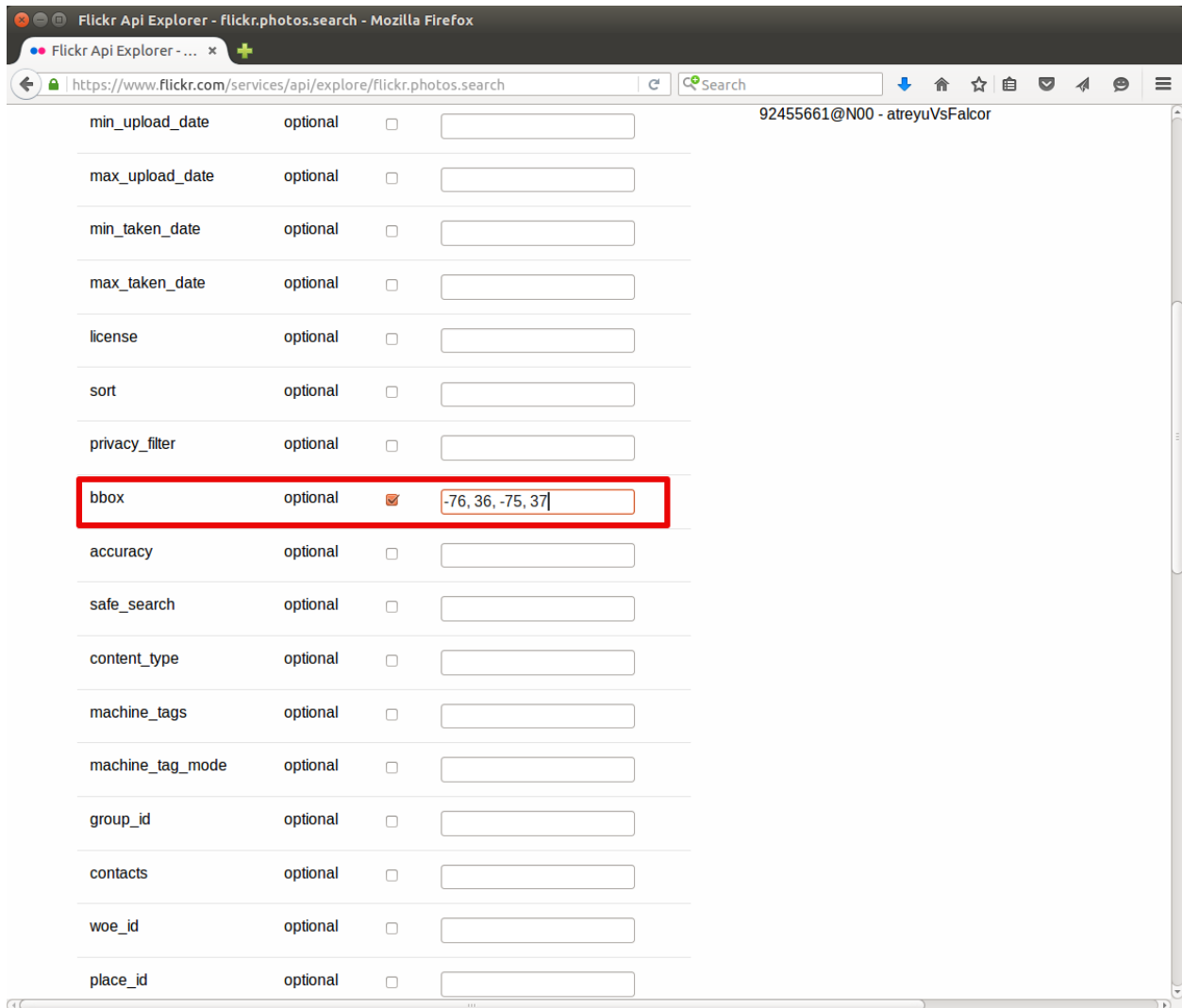


Figure 16: Flickr's API exploration page with bounding box data.

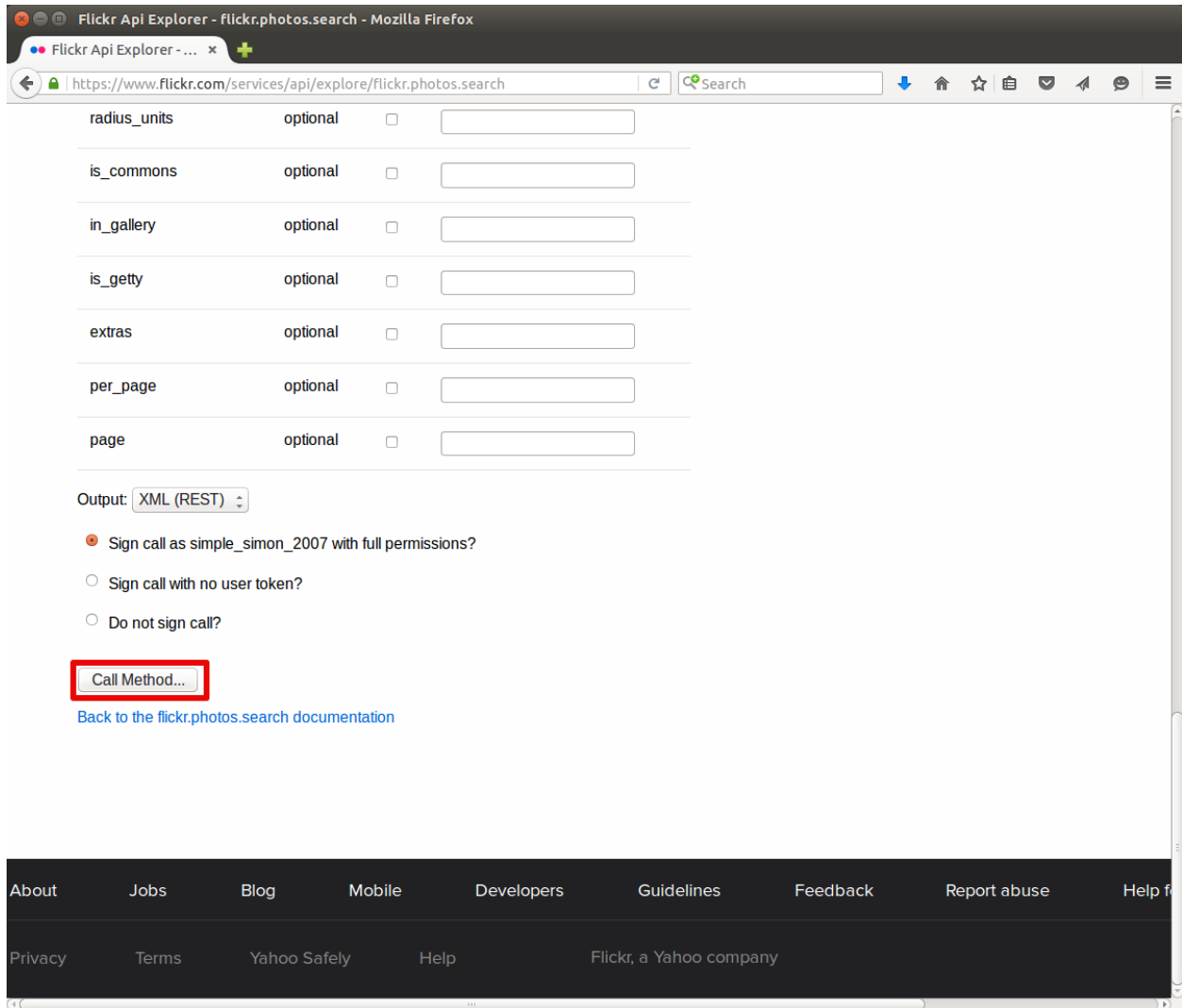


Figure 17: Flickr's API exploration calling method. The Call Method button is highlighted in red.

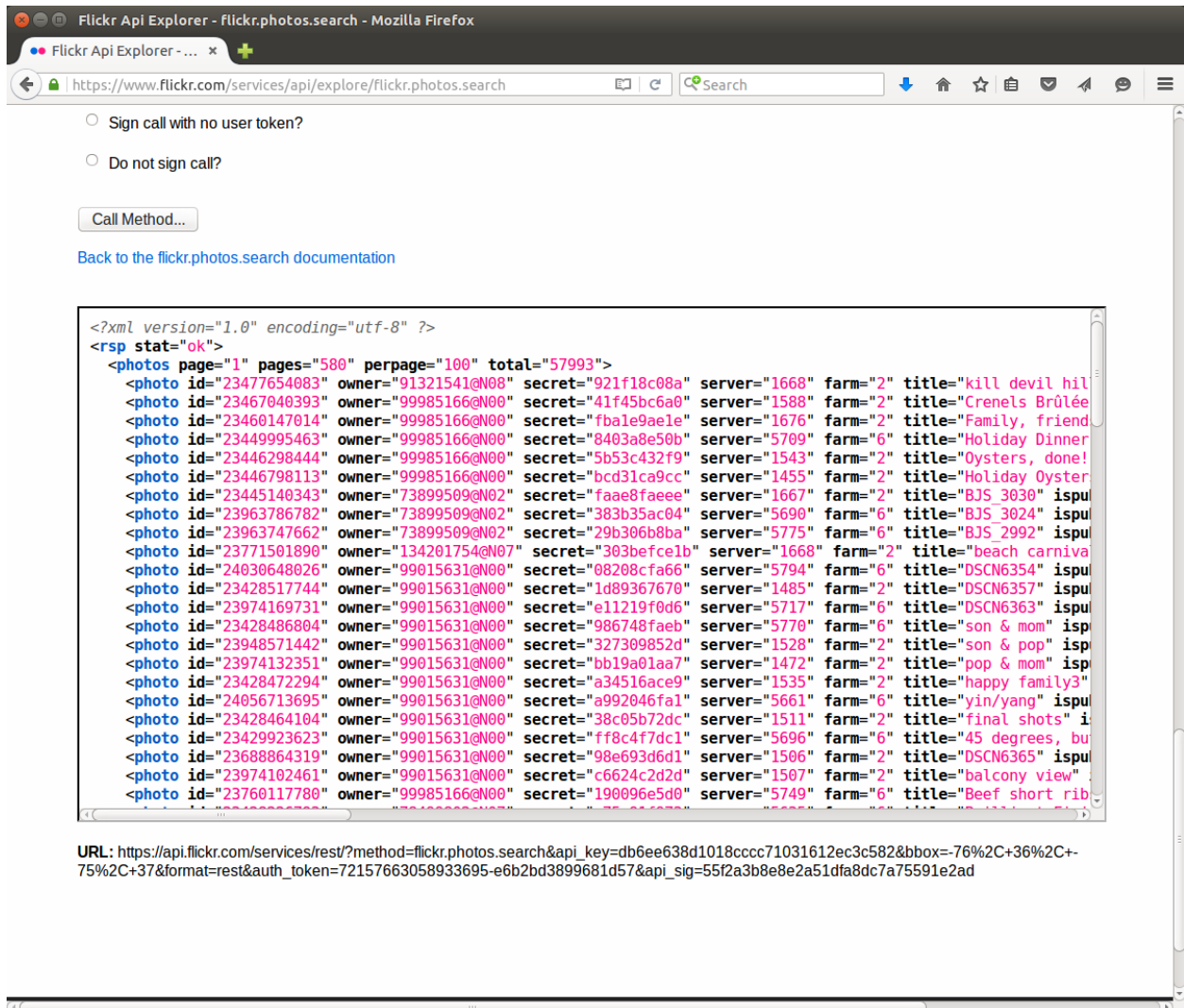


Figure 18: Flickr's API exploration with data and XML format.

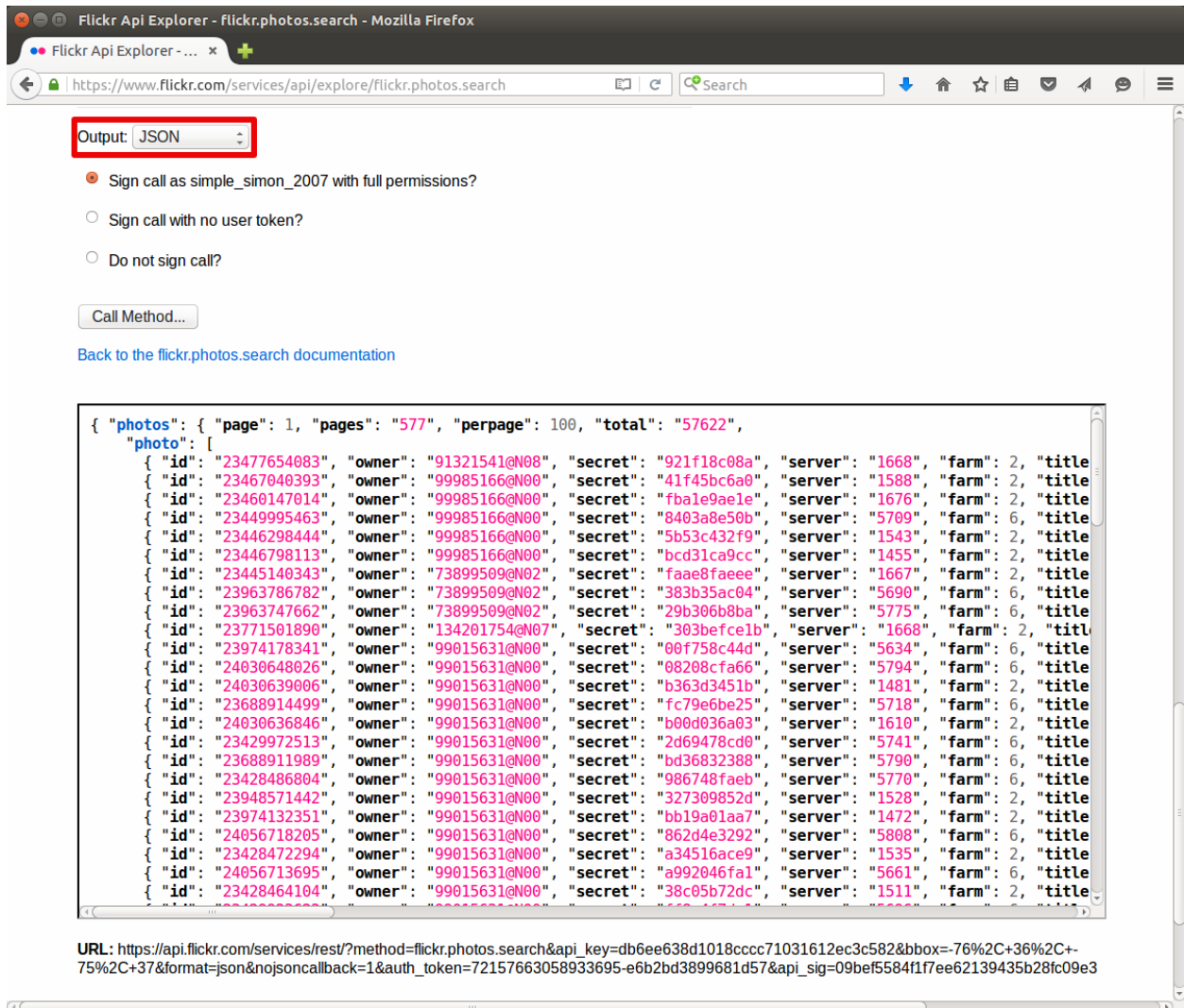


Figure 19: Flickr's API exploration with data and JSON format.

3 Selected Big Data Sources

The sites listed in this section are attached to this PDF. The R driver program used to collect the home pages for the sites in this report is attached as well. The R worker program used to do the “heavy lifting” is also attached.

3.1 Aggregator

1. 70+ websites to get large data repositories for free: A blog dealing with all things Big Data. (see Figure 20)
<http://bigdata-madesimple.com/70-websites-to-get-large-data-repositories-for-free/>

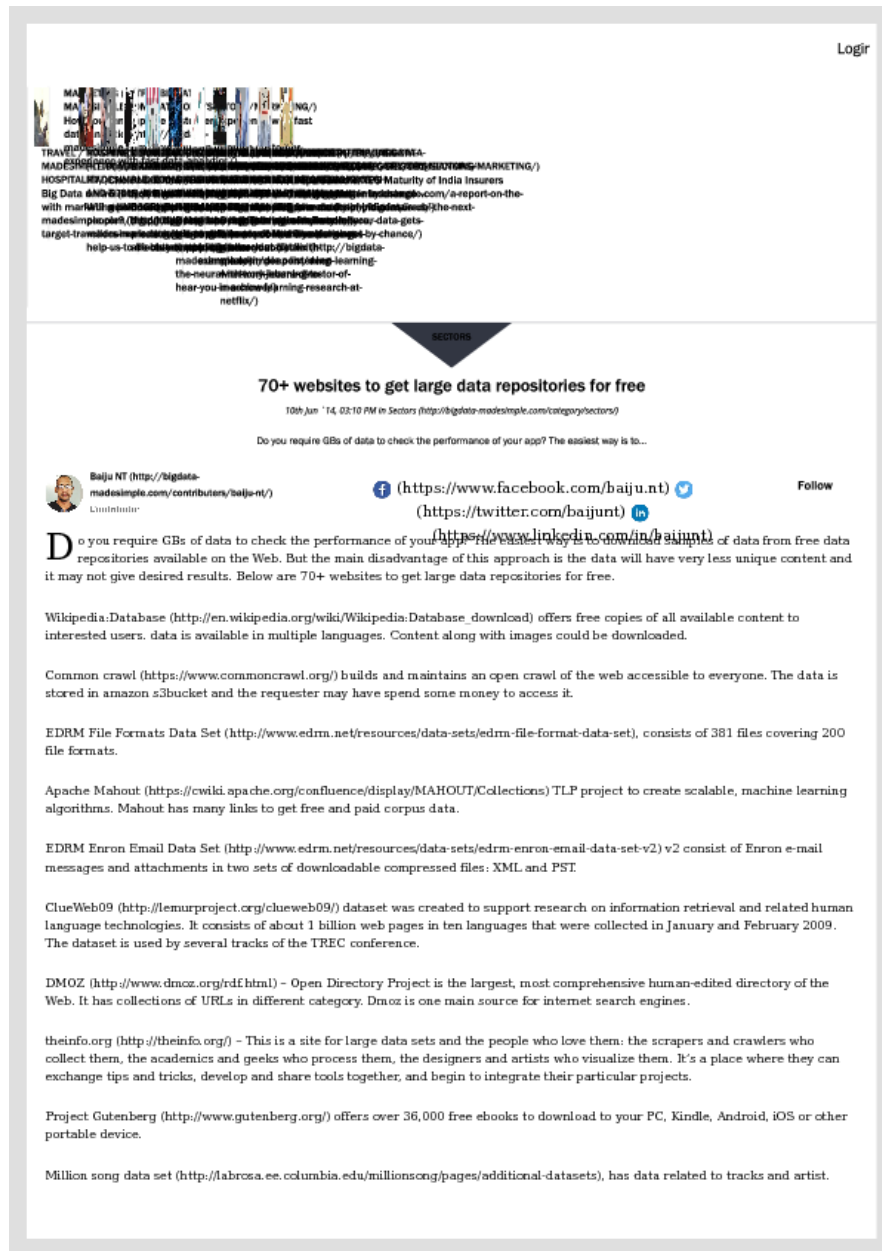


Figure 20: 70+ websites to get large data repositories for free home page.

2. Anywhere that I could download a sample database? : learnprogramming: A general question posed by many, and containing links to other sources. (see Figure 21)
https://www.reddit.com/r/learnprogramming/comments/2wr4mm/anywhere_that_i_could_download_a_sample_database/

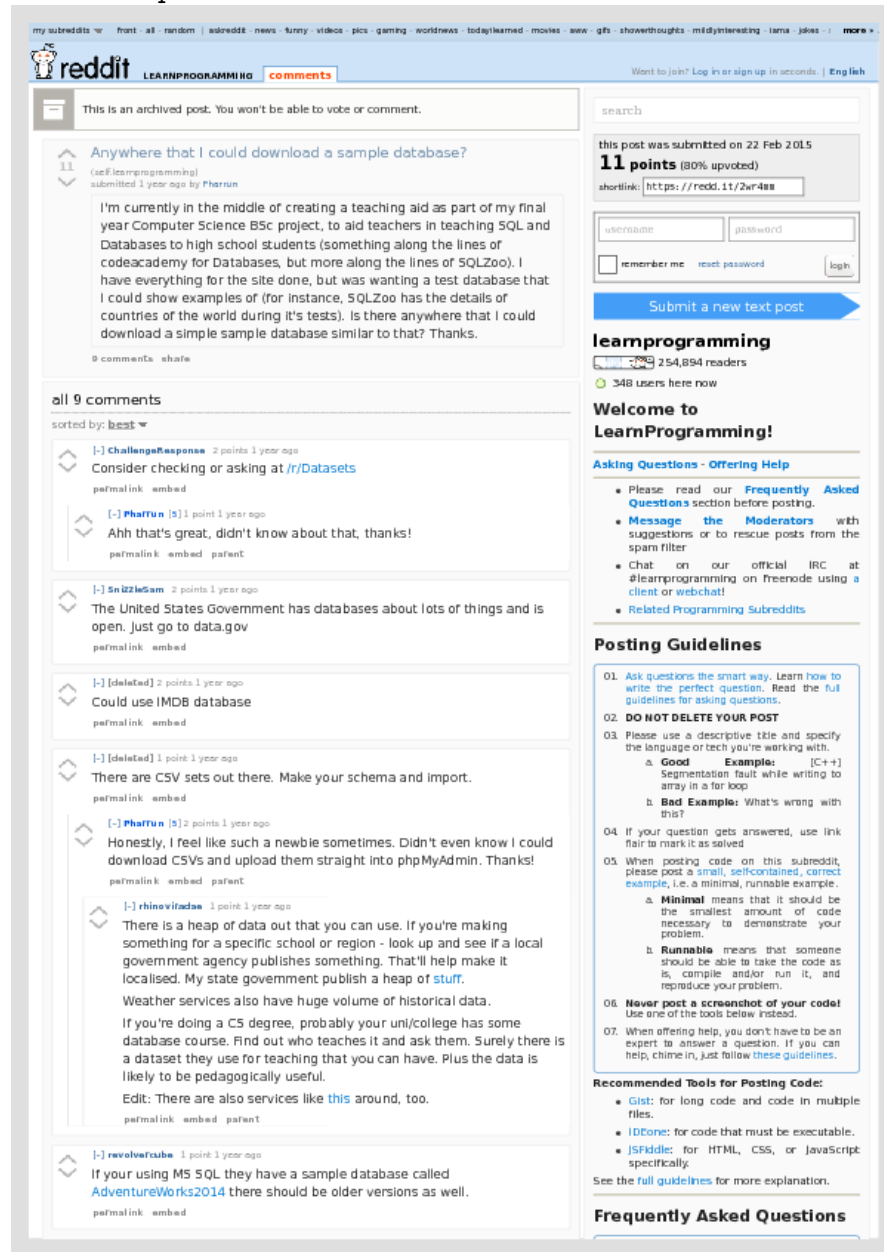


Figure 21: Anywhere that I could download a sample database? : learnprogramming home page.

3. Data Portals: A comprehensive list of open data portals from around the world. (see Figure 22)
<http://dataportals.org/>

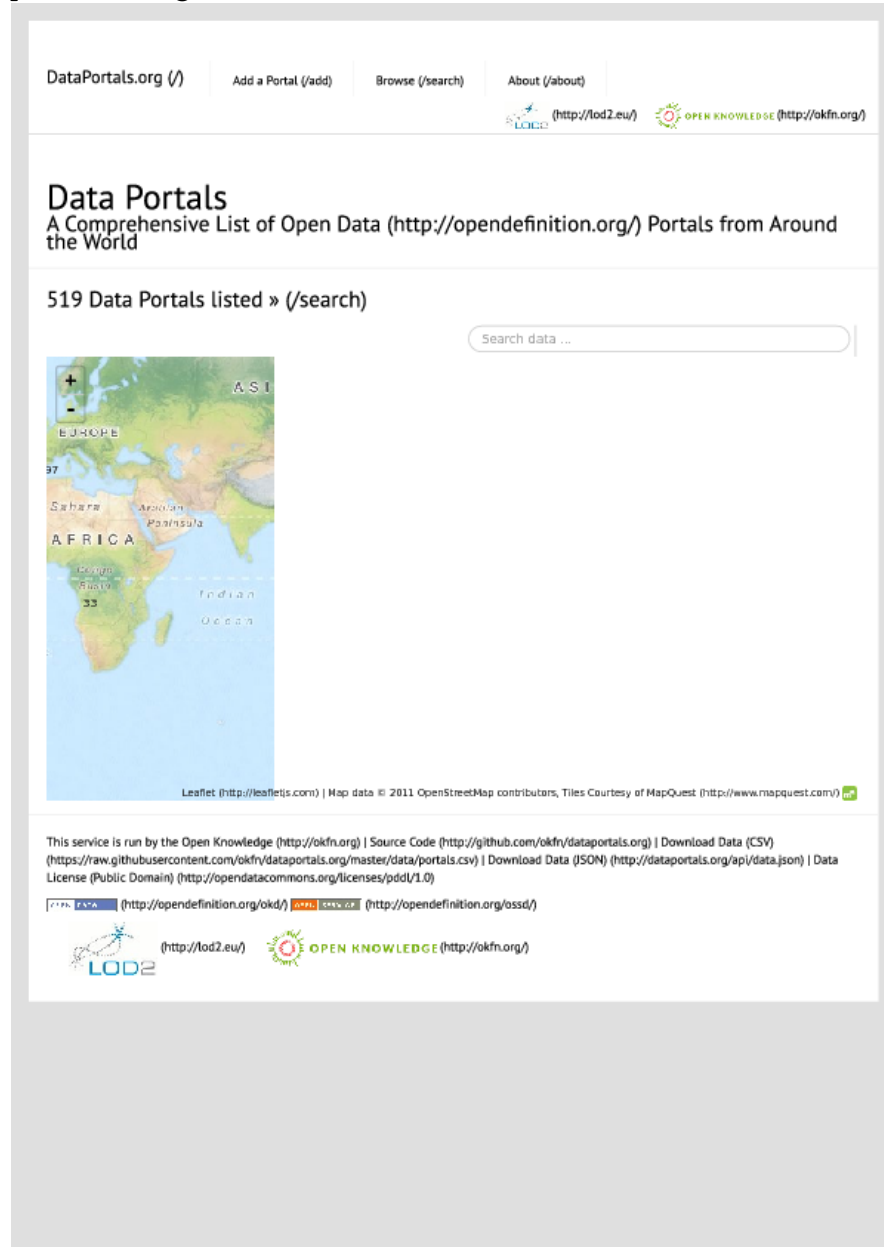


Figure 22: Data Portals home page.

4. Datasets Archive: A place to ask about the location of different datasets. (see Figure 23)
<https://www.reddit.com/r/Datasets>

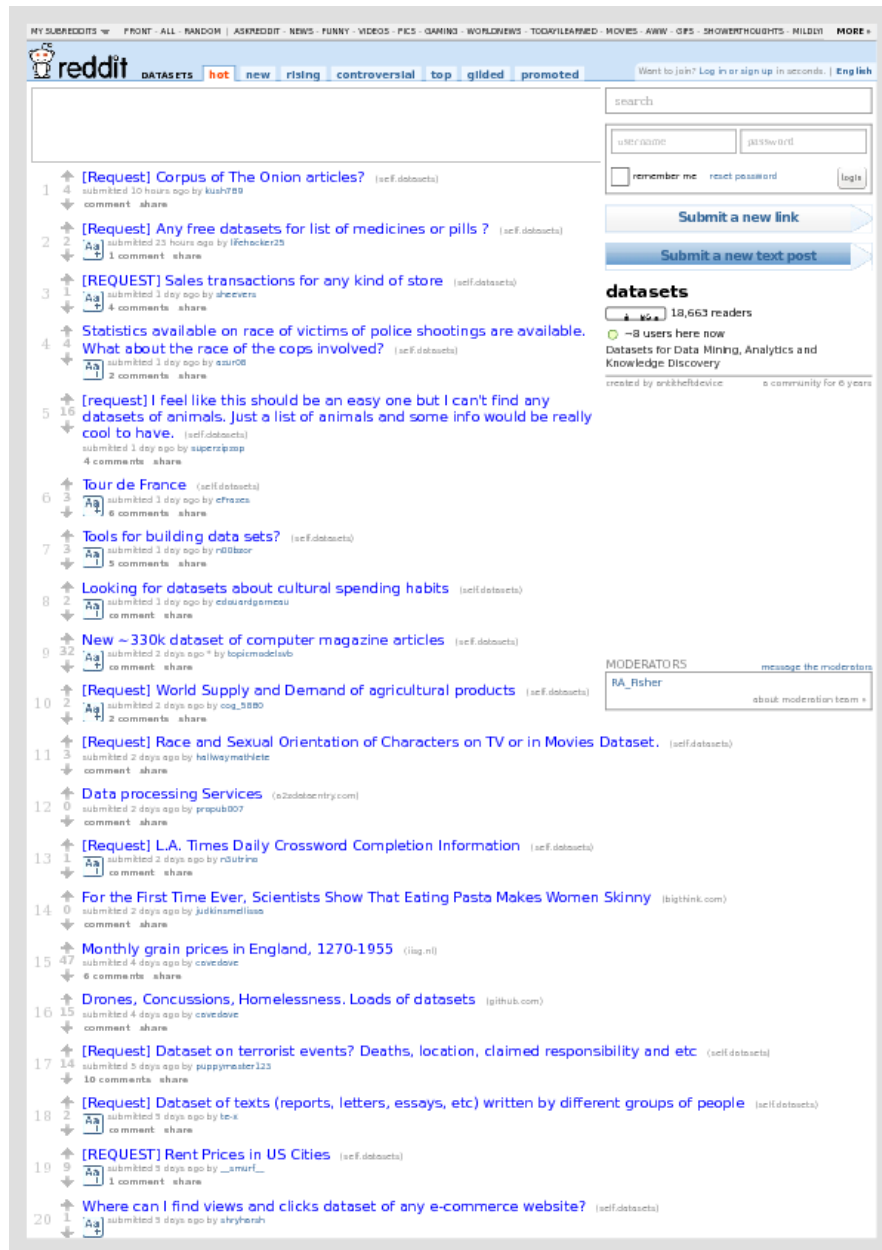


Figure 23: Datasets Archive home page.

5. Datasets for Data Mining and Data Science: A good source about data mining, analytics, big data, and data science. (see Figure 24)
<http://www.kdnuggets.com/datasets/index.html>



Figure 24: Datasets for Data Mining and Data Science home page.

6. enigma: A provider of data and data services. (see Figure 25)
<http://enigma.io/>

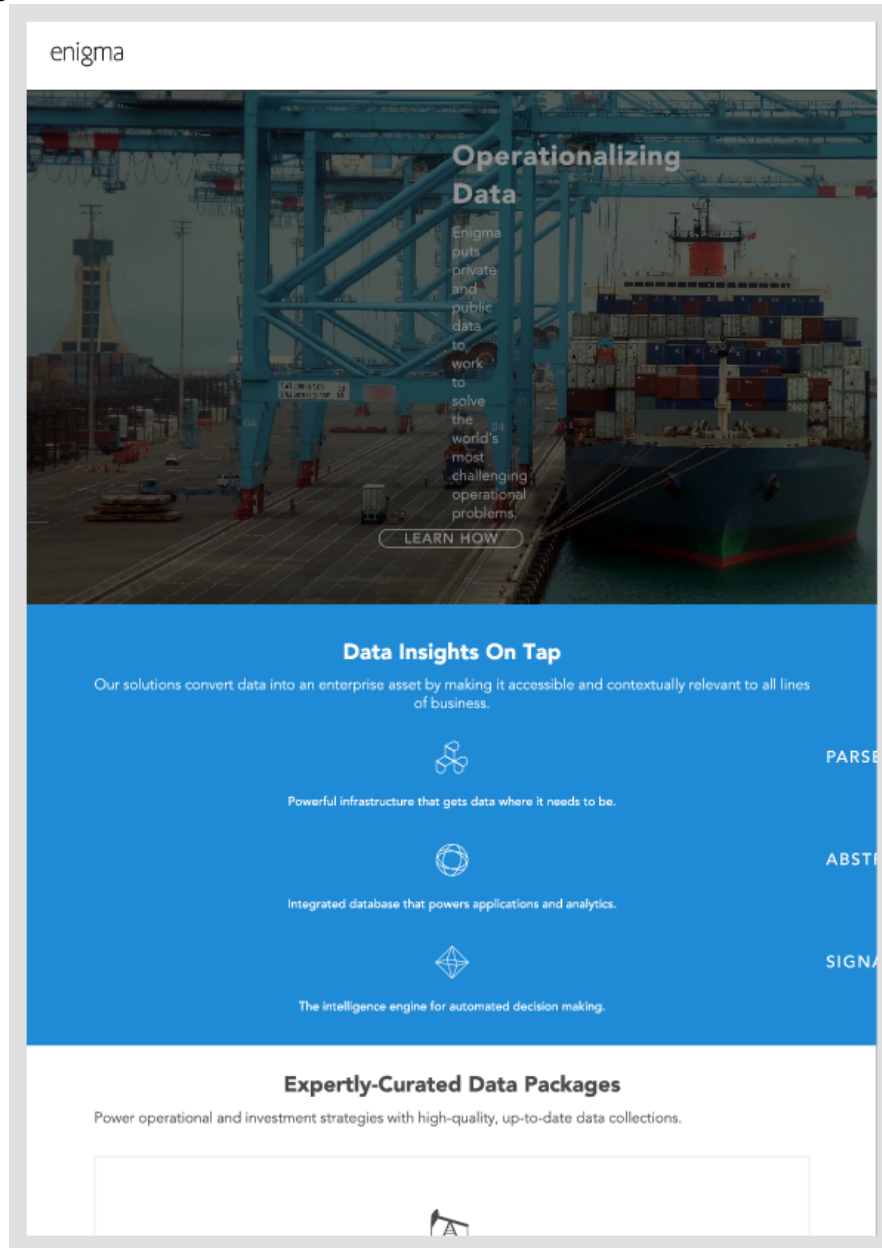


Figure 25: enigma home page.

7. Finding Data on the Internet: The following list of data sources has been modified as of 3/18/14. Most of the data sets listed below are free, however, some are not. (see Figure 26) <http://www.inside-r.org/howto/finding-data-internet>

inside-R
A Community Site for R - Sponsored by Revolution Analytics

Finding Data on the Internet

By **RevoJoe** on October 6, 2011

The following list of data sources has been modified as of 3/18/14. Most of the data sets listed below are free, however, some are not.

If an (R) appears after source this means that the data are already in R format or there exist R commands for directly importing the data from R. (See <http://www.quantmod.com/examples/intro/> for some code.) Otherwise, I have limited the list to data sources for which there is a reasonably simple process for importing csv files. What follows is a list of data sources organized into categories that are not mutually exclusive but which reflect what's out there.

Economics

American Economic Ass. (AEA): <http://www.aeaweb.org/RFE/toc.php?show=complete>
 Gapminder: <http://www.gapminder.org/data/>
 UMD: <http://inforumweb.umd.edu/econdata/econdata.html>
 World bank: <http://data.worldbank.org/indicator>

Data Science Practice

This section contains data sets used in the book "Doing Data Science" by Rachel Schutt and Cathy O'Neil (O'Reilly 2014)
 Datasets on the book site: https://github.com/oreillymedia/doing_data_science
 Enron Email Dataset: <http://www.cs.cmu.edu/~enron/>
 GetGlue (time stamped events: users rating TV shows): <http://bit.ly/1aL8X50>
 Titanic Survival Data Set: <http://bit.ly/1kj4pkF>
 Half a million Hubway rides: <http://hubwaydatachallenge.org/trip-history-data/>

Finance

CBOE Futures Exchange: <http://cfe.cboe.com/Data/>
 Google Finance: <http://www.google.com/finance> (R)
 Google Trends: <http://www.google.com/trends?q=google&ctab=0&geo=all&date=all&sort=0>
 St. Louis Fed: <http://research.stlouisfed.org/fred2/> (R)
 NASDAQ: <https://data.nasdaq.com/>
 OANDA: <http://www.oanda.com/> (R)
 Quandl: <http://www.quandl.com/>
 Yahoo Finance: <http://finance.yahoo.com/> (R)

Government

Archived national government statistics: <http://www.archive-it.org/>
 Australia: <http://www.abs.gov.au/AU55STAT5/abs/g.nsf/DetailsPage/3301.02009?OpenDocument>
 Canada: <http://www.data.gc.ca/default.asp?lang=En&n=5BCD274E-1>
 DataMarket: <http://datamarket.com/>
 FDA: <https://open.fda.gov/index.html>
 Fed Stats: <http://www.fedstats.gov/cgi-bin/A2Z.cgi>
 Guardian world governments: <http://www.guardian.co.uk/world-government-data>
 HUD: <http://www.huduser.org/portal/datasets/pdrdatas.html>
 London, U.K. data: <http://data.london.gov.uk/catalogue>
 New Zealand: http://www.stats.govt.nz/tools_and_services/tools/TableBuilder/tables-by..
 NYC data: <http://nycplatform.socrata.com/>
 OECD: http://www.oecd.org/document/0,3746,en_2649_201185_46462759_1_1_1_1_00.html

Figure 26: Finding Data on the Internet home page.

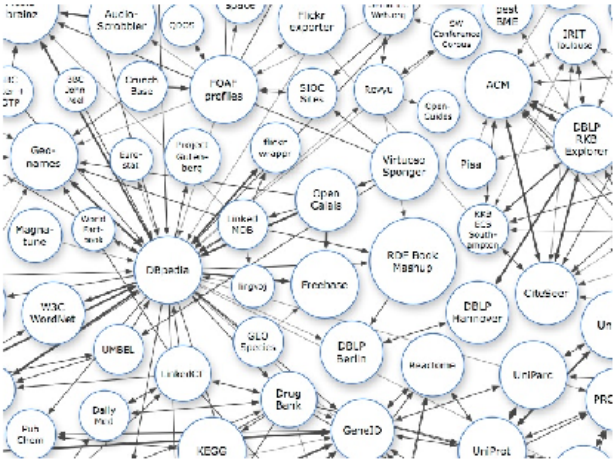
8. Linked Data - Connect Distributed Data across the Web: Linked Data is about using the Web to connect related data that was not previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. More specifically, Wikipedia defines Linked Data as "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF." (see Figure 27)
<http://linkeddata.org/>

Linked Data - Connect Distributed Data across the Web


Linked Data

Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. More specifically, Wikipedia defines Linked Data as "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF."

This site exists to provide a home for, or pointers to, resources from across the Linked Data community.



Part of the Linking Open (LOD) Data Project Cloud Diagram, click for full and historical versions...



[About this Site](#)

Figure 27: Linked Data - Connect Distributed Data across the Web home page.

9. Oxford Internet Institute University of Oxford: The Oxford Internet Institute was founded in 2001 at the University of Oxford, as an academic centre for the study of the societal implications of the Internet. (see Figure 28)

<http://geography.oi.ox.ac.uk/>



Figure 28: Oxford Internet Institute University of Oxford home page.

10. Real Time Data Resources: Many of our classroom projects use real time data sites that are difficult to find. This page contains links to some of the most compelling Internet web sites for educational use. These include a host of sites that have real time weather, air, water, and satellite information that will enhance traditional classroom lessons. (see Figure 29) <http://www.k12science.org/materials/resources/realtimedata/>

The screenshot shows the 'Real Time Data Resources' page from Stevens Institute of Technology. At the top, there are navigation links for various university departments and the CIESE logo. Below the header, the page is titled 'Real Time Data Resources' and includes a brief introductory paragraph. The main content is organized into several categories, each with a list of resource links:

- Weather and Meteorological Real Time Data Sites**
 - UH Weather (http://maui.sprl.hawaii.edu/maui/)
 - Storms: Hurricane Central (http://www.weather.com/weather/hurricane/central/)
 - GOES Satellite Images (http://rdsl.gsfc.nasa.gov/goes/) - NASA
 - The Weather Channel (http://www.weather.com/)
 - Real-Time Satellite Images and Sea Surface Temperatures (http://www.ssec.wisc.edu/data/) - University of Wisconsin
 - Oceanweather inc. (http://www.oceanweather.com/index.html) - Marine real-time data from ships and buoys
 - Weather Underground (http://www.wunderground.com/)
 - Umap Weather (http://weather.unsw.edu.au/) - Satellite and radar maps
 - Intellicast (http://www.intellicast.com/)
 - WUSA (http://www.wusa.com/) - Weather reports and alerts with clickable state map
- Astronomy And Space Science Real Time Data**
 - Recent Images from the Hubble Space Telescope (http://hubble.nasa.gov/)
 - Solar System Live (http://www.kuimlab.uchicago.edu/html/) - View the current orbital location of every planet in our solar system.
 - Space Weather Prediction Center (http://www.swpc.noaa.gov/index.html)
 - Search for Extra-Terrestrial Intelligence (SETI) (http://setiathome.berkeley.edu/)
 - Stanford Solar Center (http://lssr.stanford.edu/)
 - Earth and Moon Viewer (http://www.fourmilab.earthview.upenn.edu/)
 - Astronomy Picture of the Day (http://www.spacemedia.nasa.gov/apod/)
 - Chandrasekhar Observatory Center (Near Real Time) (http://chandra.harvard.edu/)
 - MicroObservatory (http://mo-www.harvard.edu/MicroObservatory/) - Robotic telescope that teachers and students can control and access via the net
 - Radio Meteors (http://www.spaceweather.com/glossary/radiometeor.html) - Listen to meteors entering the earth's atmosphere
- Environmental Real Time Data**
 - U.S. State Real Time Air Pollution Data (http://www.airnow.gov/)
 - EPA Watered Real Time Data by State (http://pub.epa.gov/tow/tw/realtime/index.cfm)
 - USGS Real Time Water Data (http://water.usgs.gov/realtime.html)
 - NJDEP Bureau of Air Monitoring (http://www.nj.gov/njdep/Default.htm) - Real time air quality data for New Jersey
 - Real Time data from National Sea Grant (http://seagrass.nesg.gov/)
 - My Sound Project (http://www.mysound.uconn.edu/) - Real time water quality of Long Island Sound
- Earth Science Real Time Data Sites**
 - What's New at National Oceanographic Data Center (http://www.ngdc.noaa.gov/whatsnew.html)
 - USGS Current Earthquake Information (http://quake.usgs.gov/cgi-bin/eqsdb/eqsdb.cgi) - with clickable location maps
 - USGS National Earthquake Information Center (http://earthquake.usgs.gov/earthquake/eqsdb/) - with clickable location maps
 - Volcano World (http://volcano.oregonstate.edu/)
 - Volcano Watch (http://www.ssec.wisc.edu/data/volcano.html)
- Animal Real Time Information**
 - Ark Cam (http://www.arkcam.com/)
 - Monterey Bay Aquarium Live Web Cam (http://www.montereybayaquarium.org/mba/cams/mon_mba.aspx)
 - San Diego Zoo Panda Cam (http://www.sandiegozoo.org/pandacam/index.html)
 - Sea Turtle Migration Tracking Program (http://www.ecoelab.org/sea2.html)
 - Top Five Wild Bird Cams (http://news.utsystem.edu/annals/top-5-wild-webcams-120326.html) - Discovery
 - WhaleNet Active Satellite Tags (http://whale.wheelock.edu/whalenet-stuff/tag_cover.html)
 - Cornell University Nest Cam (http://watch.birds.cornell.edu/whetcams/home/index)
- Oceanography Real Time Data**
 - United Ocean Observatory at Division Laboratory (http://hudson.stevens-tech.edu/realtime/realtime.html) - Stevens Institute of Technology
 - National Data Buoy Center (http://www.ndbc.noaa.gov/) - Real-time coastal buoy data (http://hudson.stevens-tech.edu/realtime/realtime.html)
 - Real Time Data and Images from NOAA Ships (http://www.ssec.wisc.edu/go/rt/) -
 - Oceanweather inc. (http://www.oceanweather.com/) - Marine real-time data from ships and buoys
 - Real-Time Images from Rutgers University (http://marine.rutgers.edu/rt/) -
 - Global Sea Surface Temperature (http://www.ssec.wisc.edu/rt/rtst/)
 - NOAA Interactive Data Access Retrieval System (http://www.ndbc.noaa.gov/rtst/) -
 - Real Time Tides Data (http://www.tidesandcurrents.noaa.gov/)
- Time Zone Real Time Data**
 - World Time Zone Map with Current Time (http://www.worldtimezone.com/)
 - Official U.S. Time (http://www.time.gov/)
- Archived Data Sets**
 - U.S. Census Bureau International Database (http://www.census.gov/ipeds/www/itdb.html)
 - U.S. Census Bureau - U.S. Shalester (http://www.census.gov/ipeds/www/itdb.html)
 - The Data and Story Library (http://lib.stat.cmu.edu/DASL/)
- Internet Projects Using Real-Time Data**
 - This list includes links to some of the most popular and well-supported internet projects in the math and science field.
 - CIESE Real Time Data Projects (http://www.ciese.org/materials/k12/technology/data/)
 - WhaleNet (http://whale.wheelock.edu/) - Satellite tagging
 - Journey North: A Global Study of Wildlife Migration (http://www.journey.org/north/)
 - VolcanoWorld (http://volcano.oregonstate.edu/)
 - Live From Earth and Mars (http://www.k12science.org/washington/k12/)
 - The JASON Project (http://www.jason.org/)
 - The GLOBE Program (http://www.globe.gov/)
- Other Useful Real Time Data Sites**
 - Find the distance between any two locations on the on the globe (http://www.indo.com/distance/) (http://www.math.montclair.edu/~lstarin/realtime/)
 - Yahoo! Stock Quotes (http://quote.yahoo.com/)

Figure 29: Real Time Data Resources home page.

11. Thirty Thousand Feet: Thirty Thousand Feet is an aviation directory with thousands of links to aviation web pages, aviation news, and other sources of commercial, military, and general aviation information. (see Figure 30)

<http://www.thirtythousandfeet.com/track.htm>

The screenshot shows the homepage of the Thirty Thousand Feet Aviation Directory. The main heading is "Aircraft Flight Tracking". The page is divided into several sections:

- Navigation:** HOME, AVIATION NEWS, AIRPLANE GEEKS, ADD LINK, ABOUT, CONTACT, SEARCH.
- This page:**
 - Aircraft Tracking
 - Tracking Products
 - Airport Delays
 - Radar Spotting
- See these other pages:**
 - Airlines
 - Airports
 - Air Travel
 - Travel Security Blog
 - Travel Security Links
- AirplaneGeeks.com** logo and "Follow Max on: twitter".
- OANDA** logo and a large green placeholder image.
- Related products:**
 - Airnav Live Flight Tracker for Aviation Enthusiasts: Software for Tracking Airline Flights to and From Any Airport on Your Home Computer
- Sites Offering Aircraft Flight Tracking:**
 - Airline Flight Tracker & Airport Arrivals Departures:** Track airline flights over North America, Europe, Asia, and Worldwide. Check the status of any commercial passenger flight or locate your cargo-package anywhere. See airline timetables, visit a satellite, track space flights.
 - Airwise:** Worldwide flight schedules and flight arrivals for major airlines. Look under "Flight Info."
 - Flight Arrivals:** A comprehensive and independent source of arrival and departure information for commercial airline flights over the U.S. and Canada. Also shows current airport status.
 - FlightAware:** View schedule and activity for both private and commercial air traffic in the United States. See scheduled, enroute, and recent flights for any airport. Tracking, history, maps, and graphs. Also, airport and operational information, airport diagrams, and instrument/terminal procedures for most US airports. FlightAware also maintains a user-contributed list of airport businesses such as FBOs, mechanics, and aircraft rental/flight schools along with a list of services and regularly updated fuel prices.
 - FlightExplorer.com:** This site provides real time commercial and general aviation flight tracking, airport status, and travel delay forecast information. They also have links to other useful sites: airline seat finder, fare search, weather, airport parking, WIFI hotspots, planned flight routes.
 - New! Flightradar24:** Shows live air traffic from around the world by combining data from sources including ADS-B, MLAT, and FAA, and aggregating them with schedule and flight status data from airlines and airports.
 - FlightStats:** Check world-wide flight status by flight, route, or by airport. The Flight Tracker lets you to track most commercial flights in the US and Canadian airspace on a Google Map.
 - FlightTracker.gleff.com:** A live flight tracker for aircraft transmitting ADS-B around Sydney Australia. Using an S85-1 virtual radar, aircraft are tracked and plotted on Google Maps.
 - FlightView:** Track upcoming and in-air flights on a map with radar weather, see a national airport delay map with weather overlay, FAA airport delay status, airport arrivals, departures, weather and more.
 - Flightwise:** An Internet-based tool for anyone interested in aviation. Pilots can access the web site to obtain information and plan their flights, and FBOs can use the site to track flights, attract aircraft to their facilities and help manage their operations. Charter outfits can use the site to manage their fleet. Develop your own flight tracking applications via the PlanXML Flight Tracking API.
 - Google:** Enter the two-character airline code and flight number (e.g., UA 801) and see brief flight status info on the search results page. Then click the "Track status..." link to see detailed information from FlightStats, including departure and arrival times, gates, weather, codeshare information, airport ratings, security wait times, and airline on-time performance. Also, track the flight in real-time.
 - GMaps Flight Tracker:** This site shows inbound flights for some US airports using Google Maps, as well as traces for past flights.
 - OAGflights.com:** Allows subscribers online access to worldwide flight schedules for all major airlines and low cost airlines, also including cabin class availability and arrival information.
- Aircraft Tracking Products and Services** (bottom bar)

Figure 30: Thirty Thousand Feet home page.

12. Top 10 Weather APIs: Weather has become a pretty hot topic, especially in technology circles. Weather has not only become a standard app found on nearly every smartphone and mobile device, but it's being used by many technology companies in new and innovative ways. (see Figure 31)

<http://www.programmableweb.com/news/top-10-weather-apis/analysis/2014/11/13>

ProgrammableWeb (/) Api News (category) (news)

Top 10 Weather APIs

API University (tag:university)
 Analysis (category) (news) (article) (types=analysis), Weather (category) (weather), Climate (category) (climate)
 Nov 13 2014
 By James Wagner (author) (wagnerj, ProgrammableWeb Staff)

Weather has become a pretty hot topic, especially in technology circles. Weather has not only become a standard app found on nearly every smartphone and mobile device, but it's being used by many technology companies in new and innovative ways. (see Figure 31)

Open this page on a mobile device to see how you can use weather data to better personalize your website and create weather-informed advertising. In June, the Weather Channel announced (<http://weather.com/2014/06/11/twitter-weather-ads/>) a partnership with Twitter to enable weather-informed advertising on the Twitter website and mobile applications. Advertisers will be able to create promoted tweets that target specific audiences based on weather data. A few months ago, Zappos, a popular Amazon.com-owned online shoe and clothing shop, partnered (<http://www.venturebeat.com/2014/06/09/zappos-uses-third-party-weather-data-better-personalized-with-choose2share-to-deliver-weather-based-personalized-advertising-to-zappos-customers-choose2share-uses-a-weather-forecast-api-to-create-advertisements-that-consider-customers-local-weather-conditions-in-real-time/>) with Choose2Share to deliver weather-based, personalized advertising to Zappos customers. Choose2Share uses a weather forecast API to create advertisements that consider customers' local weather conditions in real time.



BloomSky is a five-in-one weather station and HD camera that captures real-time weather data along with sky images at your location. The goal of BloomSky is to build a crowdsourced meteorological network. Another recent and increasingly popular trend in the crowdsourcing of weather-related applications and devices. Examples include:

- BloomSky** (<http://www.bloomsky.com/>) — Described by the project as the “world’s first smart weather camera,” BloomSky is a weather probe that consumers can place in a yard or on a roof that captures hyperlocal weather conditions data. The goal of BloomSky is to build a crowdsourced meteorological network that provides real-time weather monitoring and weather data to consumers.
- Climate** (<http://www.kickstarter.com/projects/662620533/climate-create-your-own-friendly-environment>) — Climate is a Bluetooth-enabled environment tracker that measures humidity, temperature and UV. The data is synced to the consumer’s smartphone, and each piece of data is used to power WeatherBlox, a crowdsourced, real-time weather map.
- Metab** (<http://metab.com/>) — Metab is a crowdsourced API platform that provides local weather and environmental data to third-party applications. Metab collects weather information from Twitter, Instagram, the National Oceanic and Atmospheric Administration and other sources.
- Wexzo** (<http://www.wexzo.com/>) — Wexzo is a crowdsourced weather app available on Android and iOS. Users can share local weather observations and access real-time weather maps from around the world.
- Weathermob** (<http://www.weathermob.me/>) — Weathermob is a large community-based weather service that captures real-time weather data via mobile, social and meteorological data sources. Weathermob is only available on iOS.
- StormTag** (<http://openignal.com/blog/2014/06/13/stormtag-fully-funded/#WeatherSignal>) (<http://weathersignal.com/>) — StormTag and WeatherSignal are weather-based projects created by OpenSignal. StormTag is a Bluetooth weather station that attaches to a keychain. The weather data collected by the device is sent to the WeatherSignal app and OpenSignal servers.

BloomSky, Climate and StormTag/WeatherSignal are not only crowdsourced but were all successfully funded via Kickstarter campaigns. At the time of this writing, there were more than 100 weather-related Kickstarter campaigns (<https://www.kickstarter.com/>), many of which are in the technology and wearable categories.



Climate is a Bluetooth-enabled environment tracker that measures humidity, temperature and UV. The Climate project was successfully funded through Kickstarter. Image credit: Climate (<http://www.kickstarter.com/projects/662620533/climate-create-your-own-friendly-environment/>)

The increasing prevalence of smartphones and mobile devices, along with the exponential rise of the Internet of Things, has helped fuel the interest in and popularity of crowdsourced weather applications and devices. All of these innovative weather-related devices and applications would not have been possible without the advent of APIs, specifically weather APIs.

Related: Top Weather APIs: an In-Depth Comparison (http://www.programmableweb.com/news/top-weather-apis-in-depth-comparison/analysis/2015/02/12?utm_source=edit&utm_medium=partners&utm_campaign=related-content)

This article features ProgrammableWeb’s Top 10 Weather APIs, which have been selected based on a variety of metrics, including the number of FV followers, GitHub activity, Twitter activity and recent news. The APIs are listed in alphabetical order.

ProgrammableWeb’s Top 10 Weather APIs

AccuWeather

AccuWeather (<http://www.accuweather.com/>) is one of the leading digital weather information providers. According to its website, AccuWeather provides weather forecasts for nearly 5 million locations worldwide, and over a billion people worldwide rely on AccuWeather every day.

The company has been working steadily toward broadening its audience as well as adding new weather-related products and services. Late last year, the company acquired (<http://www.prmwires.com/news-releases/accuweather-launches-its-new-weather-forecast-app-for-ios-and-android-231068533.html>) Sky Station, a developer of extreme short-term and highly localized weather forecasts. The acquisition of Sky Station allowed AccuWeather to launch (<http://www.prmwires.com/news-releases/accuweather-launches-its-new-weather-forecast-on-web-and-in-app-for-iphone-ipod-touch-240448553.html>) its MinuteCast feature, which was introduced earlier this year. In February, AccuWeather announced (<http://www.prmwires.com/news-releases/accuweather-provides-forecasts-for-sms-using-newly-acquired-gdxy-e5-247482443.html>) that it had partnered with Samsung to provide a pre-installed weather app on the Samsung Galaxy S5 smartphones. The company even has plans to launch the AccuWeather Channel, a 24/7 weather channel.

Developers can use the AccuWeather API (<http://www.programmableweb.com/api/accuweather>) to add severe weather alerts, daily and hourly forecasts, current conditions, and other weather-related functions to their applications. No freemium is plan available; AccuWeather only offers paid plans. Developers interested in using the AccuWeather API can contact the company’s sales department for pricing.

Australian Bureau of Meteorology

The Australian Bureau of Meteorology (<http://www.bom.gov.au/>) is the government agency responsible for providing Australians with assistance in dealing with drought, flood, fire, storms and other harsh weather conditions. The agency also provides a variety of weather, climate and water data, including real-time observations, computer model forecasts and agricultural bulletins.

In March, the bureau launched (<http://www.bom.gov.au/local/stories/2014/03/25/971272.html>) the Weather Observation Website (WOW), a project created in partnership with the United Kingdom Met Office. The WOW Project (<http://bom-weatheroffice.gov.uk/>) is another example of weather crowdsourcing. Australian weather enthusiasts can share weather observations, photos and other information using the WOW website (the WOW website is free to use).

Figure 31: Top 10 Weather APIs home page.

13. University Corporation for Atmospheric Research: Search and access 178 data sets covering the Atmosphere, Ocean, Land and more. Explore climate indices, reanalyses and satellite data and understand their application to climate model metrics. This is the only data portal that combines data discovery, metadata, figures and world-class expertise on the strengths, limitations and applications of climate data. (see Figure 32)
<https://climatedataguide.ucar.edu/>



Figure 32: University Corporation for Atmospheric Research home page.

14. Where can I find large datasets open to the public? - Quora: The best answer to any question, especially about where to find Big Data. (see Figure 33)
<http://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public>

The image shows a screenshot of a Quora page. At the top, there is a search bar with the text 'Search for questions, people, and topics' and a 'Sign In' button. Below the search bar, there are navigation tabs for 'Datasets', 'Data Science', 'Big Data', and 'Seeking Question'. The main heading of the page is 'Where can I find large datasets open to the public?'. Below this heading, there is a promotional banner for Udacity.com with the text 'Develop new tech skills. Be in demand. Become a Data Analyst in this self-paced course. Job offer guaranteed, or your money back.' and a 'Learn More at Udacity.com' button. The main content of the page is an 'Answer Wiki' section that lists various data repositories and search engines. The list includes 19 items, each with a URL and a small icon. The first item is 'https://www.kaggle.com/datasets' and the last is 'http://data.sigov.org'. There is also a section titled 'Single datasets and data repositories' with 10 items, each with a URL and a small icon. On the right side of the page, there is a 'Related Questions' section with several questions, each with a small icon. The questions include 'Where can I find large datasets closed to the public?', 'Where can I find large datasets open to the public for India specifically?', 'Have a link to a large free e-mail dataset (not Enron)?', 'What are some free but large datasets of general products?', 'Where can I find datasets (open to public) of eCommerce websites?', 'Where can I find large data sets open to the public of all available drugs and medicines?', 'Where can I get public spatial datasets?', 'Where can I find large datasets open to the public for merger and acquisition integration performance?', 'What large, open and public datasets are there for Educational Data Mining?', and 'Where can I find large bank and credit related datasets open to the public?'.

Figure 33: Where can I find large datasets open to the public? - Quora home page.

15. Wikipedia:Database download: Wikipedia offers free copies of all available content to interested users. These databases can be used for mirroring, personal use, informal backups, offline use or database queries (such as for Wikipedia:Maintenance). All text content is multi-licensed under the Creative Commons Attribution-ShareAlike 3.0 License (CC-BY-SA) and the GNU Free Documentation License (GFDL). Images and other files are available under different terms, as detailed on their description pages. (see Figure 34)
https://en.wikipedia.org/wiki/Wikipedia:Database_download

Wikipedia:Database download

From Wikipedia, the free encyclopedia
For scheduling, related tools etc., see m:Data dumps.
"WP:DD" redirects here. For Duplication detector, see Wikipedia:Duplication detector.

Wikipedia offers free copies of all available content to interested users. These databases can be used for mirroring, personal use, informal backups, offline use or database queries (such as for Wikipedia:Maintenance). All text content is multi-licensed under the Creative Commons Attribution-ShareAlike 3.0 License (CC-BY-SA) and the GNU Free Documentation License (GFDL). Images and other files are available under different terms, as detailed on their description pages. For our advice about complying with these licenses, see Wikipedia:Copyrights.

Contents

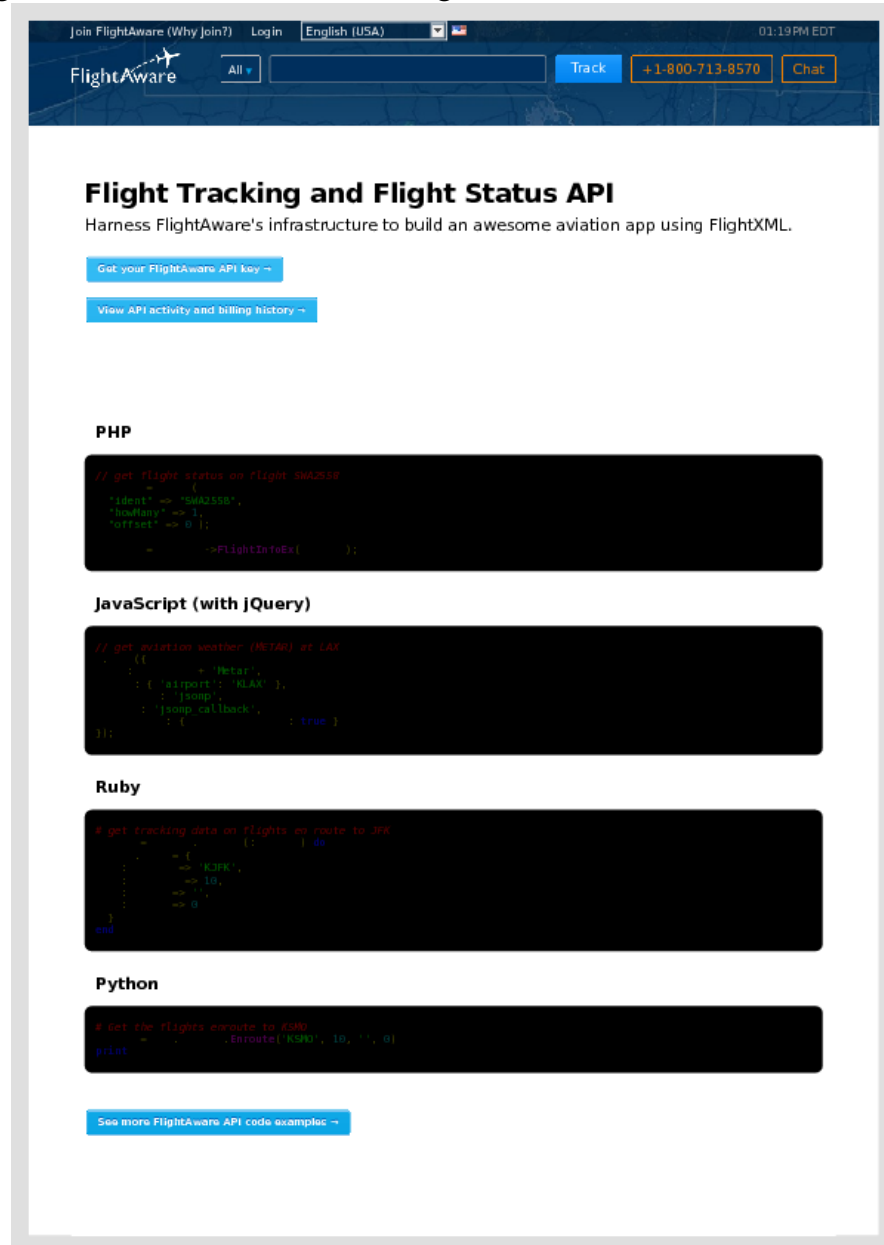
- 1 Offline Wikipedia readers
- 2 Where do I get...
 - 2.1 English-language Wikipedia
 - 2.2 Other languages
- 3 Where are the uploaded files (image, audio, video, etc., files)?
- 4 Dealing with compressed files
- 5 Dealing with large files
 - 5.1 File system limits
 - 5.2 Operating system limits
 - 5.3 Tips
 - 5.3.1 Detect corrupted files
 - 5.3.2 Reformatting external USB drives
 - 5.3.3 Linux and Unix
- 6 Why not just retrieve data from wikipedia.org at runtime?
 - 6.1 Please do not use a web crawler
 - 6.1.1 Sample blocked crawler email
 - 6.2 Doing SQL queries on the current database dump
- 7 Database schema
 - 7.1 SQL schema
 - 7.2 XML schema
- 8 Help to parse dumps for use in scripts
 - 8.1 Doing Hadoop MapReduce on the Wikipedia current database dump
- 9 Help to import dumps into MySQL
- 10 Static HTML tree dumps for mirroring or CD distribution
 - 10.1 Kiwix
 - 10.2 Aard Dictionary
 - 10.3 E-book
 - 10.4 Wikiviewer for Rockbox
 - 10.5 Old dumps
- 11 Dynamic HTML generation from a local XML database dump
 - 11.1 XOWA

Figure 34: Wikipedia:Database download home page.

3.2 Aviation

1. Flight Aware: Harness FlightAware's infrastructure to build an awesome aviation app using FlightXML. (see Figure 35)

<http://flightaware.com/commercial/flightxml/>



Join FlightAware (Why Join?) Login English (USA) 01:19 PM EDT

FlightAware All Track +1-800-713-8570 Chat

Flight Tracking and Flight Status API

Harness FlightAware's infrastructure to build an awesome aviation app using FlightXML.

[Get your FlightAware API key →](#)

[View API activity and billing history →](#)

PHP

```
// get flight status on flight 006258
<?php
    $ident => "006258",
    $rowId => 1,
    $offset => 0 );
    =>FlightInfoEx] );
```

JavaScript (with JQuery)

```
// get aviation weather (METAR) at LAX
$.ajax({
  url: "/metar",
  dataType: "json",
  success: function(data) {
    console.log(data);
  }
});
```

Ruby

```
# get tracking data on flights en route to JFK
require 'flightaware'
client = FlightAware::Client.new('API_KEY')
client.get_tracking_data('KJFK', 10, 0)
```

Python

```
# get the flights enroute to KJFK
from flightaware import FlightAware
client = FlightAware('API_KEY')
client.get_tracking_data('KJFK', 10, 0)
```

[See more FlightAware API code examples →](#)

Figure 35: Flight Aware home page.

2. Flight Stats: A collection of flight related information including arrival, departures, and current status (see Figure 36)
<https://developer.flightstats.com/>

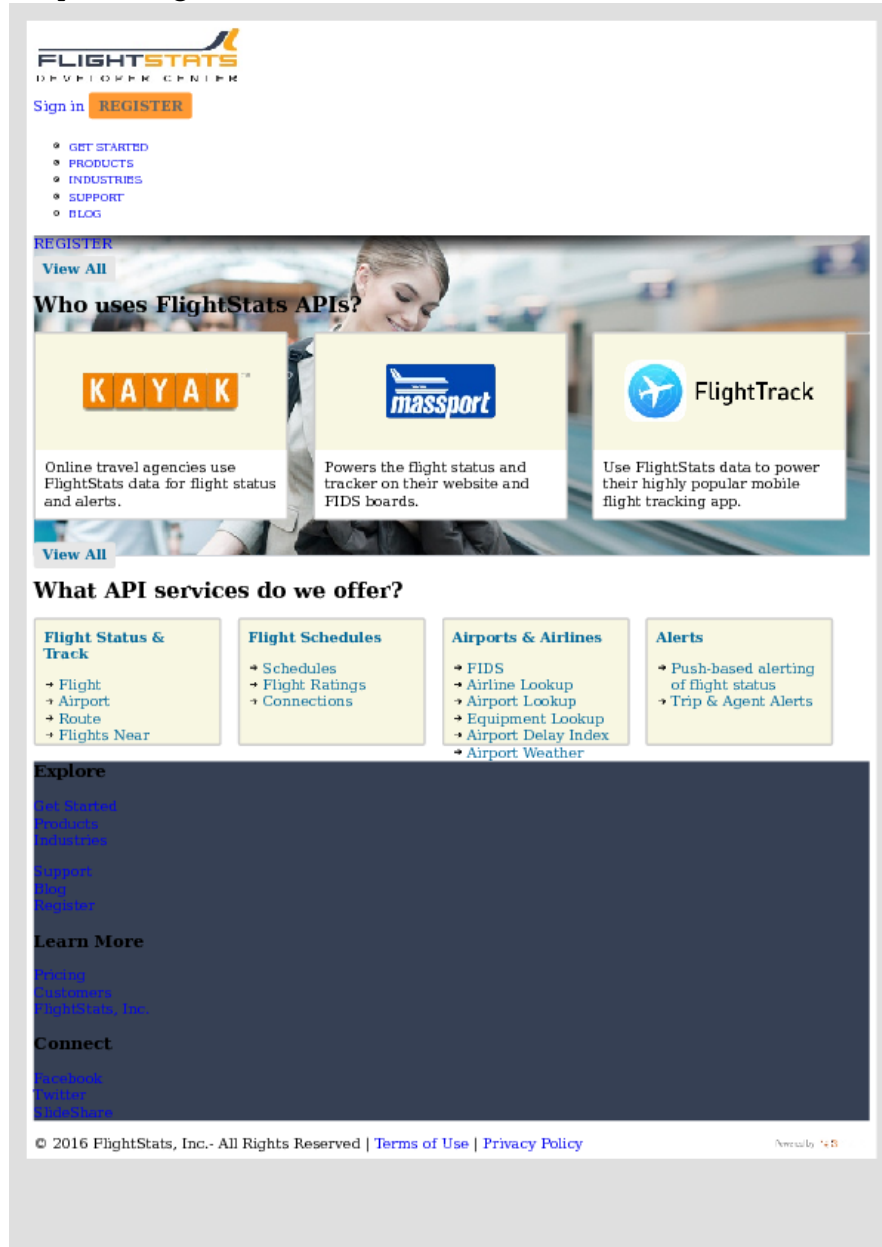


Figure 36: Flight Stats home page.

3. Flight View: FlightViews comprehensive, fast, accurate global real-time flight information aggregates data from hundreds of sources on more than 130,000 daily flights into a single truth data stream. Data customers and developers access FlightViews flight information through flexible and easily consumed APIs to power displays and applications that inform travelers and the businesses that provide travel services. (see Figure 37)
<http://corporate.flightview.com/products/data-feeds-apis>

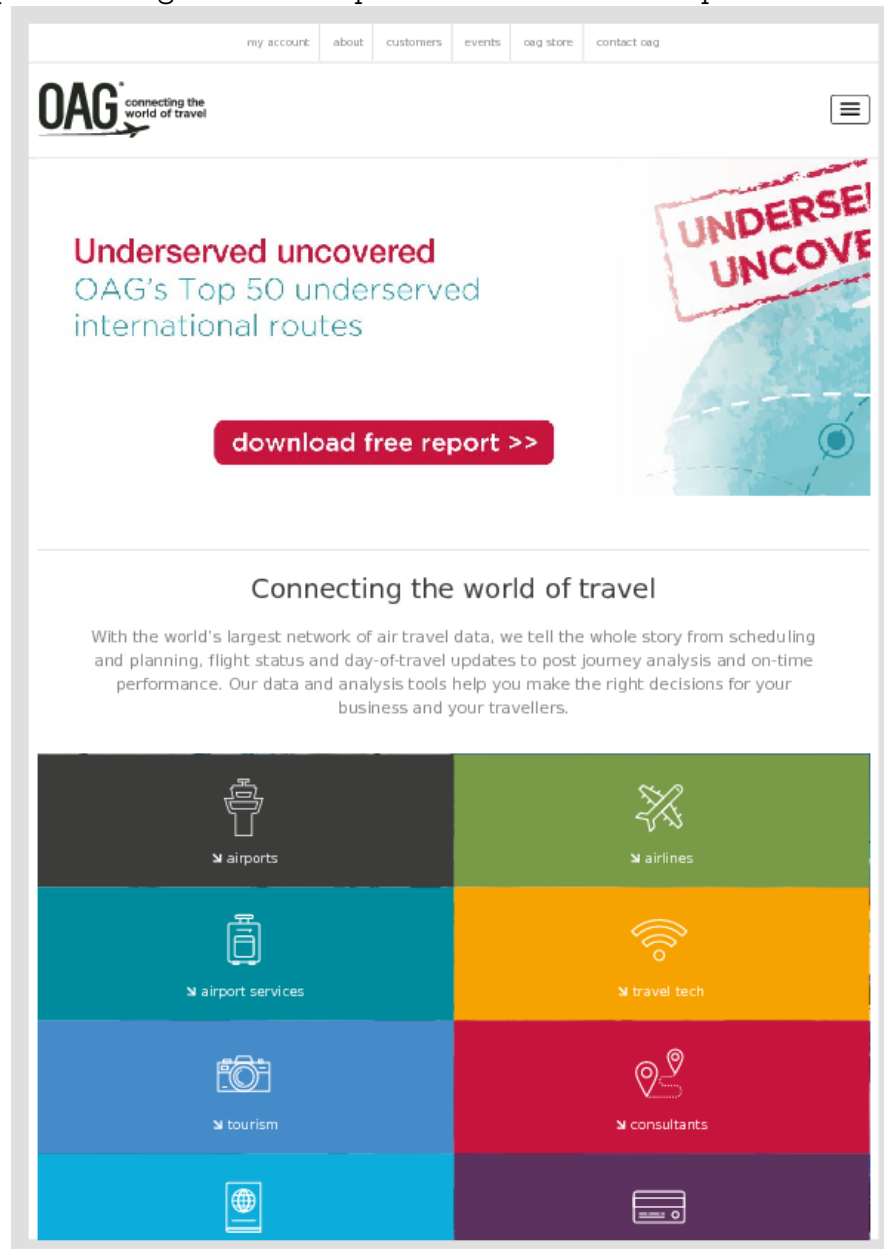



Figure 37: Flight View home page.

4. OAG (Official Airline Guide): OAG has the knowledge, experience and capability to provide reliable, high quality, up-to-the-minute aviation data, offering essential information and innovative travel planning solutions. Our products are vital to strategic and commercial planning, driving key business decisions and delivering Absolute Aviation Advantage across the worldwide air transport industry. (see Figure 38)
<http://www.oag.com/Flight-Schedules/Schedules-OnDemand>

my account | about | customers | events | oag store | contact oag

OAG connecting the world of travel

schedules ondemand API

 Power your website and other applications with a simple, fast API and access accurate flight schedules information when you need it.

With the schedules ondemand API, you can control your own supply of airline schedules and ensure you have the most up-to-date information you need.

Aggregating and normalising data costs time and money - avoid those operating costs by using an easy and accessible API that can do all the hard work for you. With access to flights other aggregators simply don't have, you will have a distinct advantage in the marketplace.

To learn more, [download the schedules on demand API product sheet](#) now

Request a free trial

Simply fill in your details below and we will be in touch to arrange your trial.

First Name*

Last Name*

Company Name*

Job Title*

Email*

Phone Number*

Industry Sector (Choose...)*

Country (Choose...)*

Please provide any comments on your business needs

[make request](#)

Take control of your airline schedules data supply with our premier web service

Key features

Figure 38: OAG (Official Airline Guide) home page.

5. QPX Express API: Global airline pricing and shopping in a single, standard API. (see Figure 39)

<https://developers.google.com/qpx-express/>

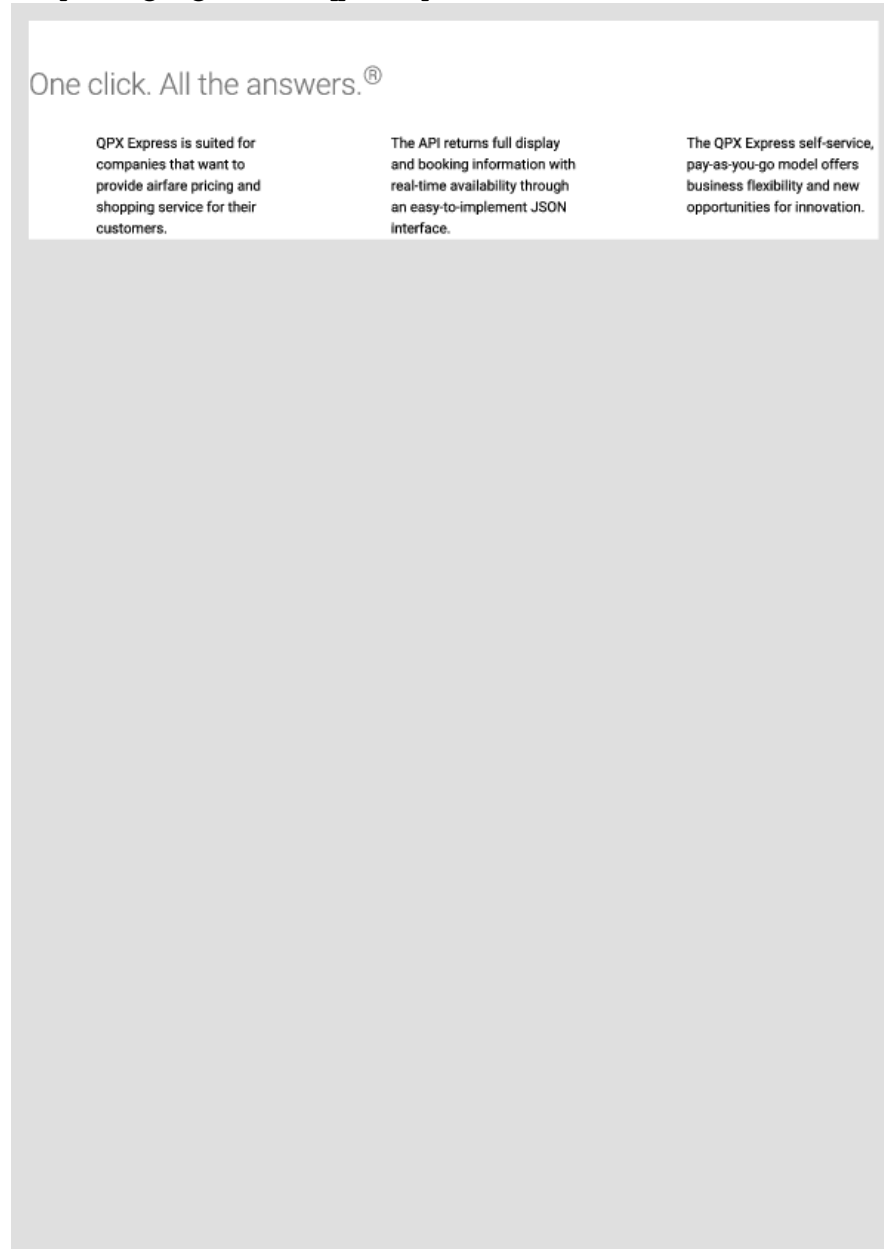


Figure 39: QPX Express API home page.

6. Travel Boutique: In recent years, the flight booking segment of the travel sector has become increasingly competitive. With the largest tour operators and websites offering customers a wide range of options all in one place, it's difficult for the smaller travel agents to compete. As ever, at Travel Boutique Online, we've developed a solution which allows small- and medium-sized travel agents in India to overcome this challenge. This solution will enable you to sell a vast inventory of flights from a large pool of airlines direct to your customers. (see Figure 40)

http://www.travelboutiqueonline.com/Flight_api.aspx

The screenshot shows the 'Flight XML API' page on the Travel Boutique Online website. The page layout includes a top navigation bar with the logo and a 24x7 helpdesk number. Below this is a banner image featuring the text 'XML API' and an airplane. The main content area is divided into a left sidebar with a navigation menu and a main content column. The sidebar menu includes 'XML/API', 'HOTELIERS', 'ONLINE PACKAGES', 'SUPPLIERS', 'TRAVEL INSURANCE', and 'Contact Us'. The main content column is titled 'Flight XML API' and contains the following sections:

- XML/API**: A sub-menu with 'Hotel XML API' and 'Flight XML API'.
- WHITE LABEL SOLUTIONS**: A section with a list of services including 'HOTELIERS', 'ONLINE PACKAGES', 'SUPPLIERS', and 'TRAVEL INSURANCE'.
- Contact Us**: A form with fields for 'Your Name*', 'Email', 'Phone No.', 'Category', 'Our Products', 'City', and 'Your Query', along with a 'Enter the Captcha' field and 'Submit' and 'Reset' buttons.
- Flight XML API**: A section with a description of the service and a list of benefits:
 - Flight search & domestic and international
 - Easy implementation
 - Integration support
 - Real time booking
 - Ticket cancellation, reissuance & real time cancellation of LCCs
- What is an API?**: A section explaining that API stands for Application Programming Interface and how it allows access to services.
- USPs**: A section titled 'USPs' (Unique Selling Points) listing:
 - New airline sources recently added include AirCosta, AirAsia, FlyDubai and AirIndiaExpress
 - Special Service Requests (SSR) have been incorporated, such as additional baggage allowance requests and meal specifications
 - Air fare calendar
 - Multi-way search
- Benefits**: A section titled 'Benefits' listing:
 - Once the flight API has been integrated with your website, you will immediately start to discover the great advantages it offers to your business.
 - Vast Inventory**: With the flight details of numerous airlines aggregated and integrated into a single search and booking engine, the inventory you are able to offer to your customers will increase dramatically.
 - Time saving**: Booking flight tickets will become a much quicker, slicker process, due to a user-friendly interface and instant confirmation. No longer will you need to enter data manually on multiple occasions, as our API will hold all the details and information you need in one place.
 - Cost savings**: Cost savings will be evident in many areas. Software is not needed, so no installation or maintenance will be required. The ability to book instantly, with real time pricing and availability on display will save a great deal of time and reduce operational costs. Integration with the TBO flight API will give you access to all the airlines we've partnered with for a one-time fee. This means

Figure 40: Travel Boutique home page.

7. Travel Fusion: TF.Flight API is the largest direct connect LCC and (a growing number of) FSC airline carriers platform available from a single API. TF.Flight API aggregates information from hundreds of airlines, providing agents, e-commerce platforms, search and mobile services with a single system to manage, search, book and support this vast content. TF.Flight API is available via an API, any Agency Desktop, and web-based (agency/user login) portals. (see Figure 41)
<https://www.travelfusion.com/corporate/page/tf-flight-api>

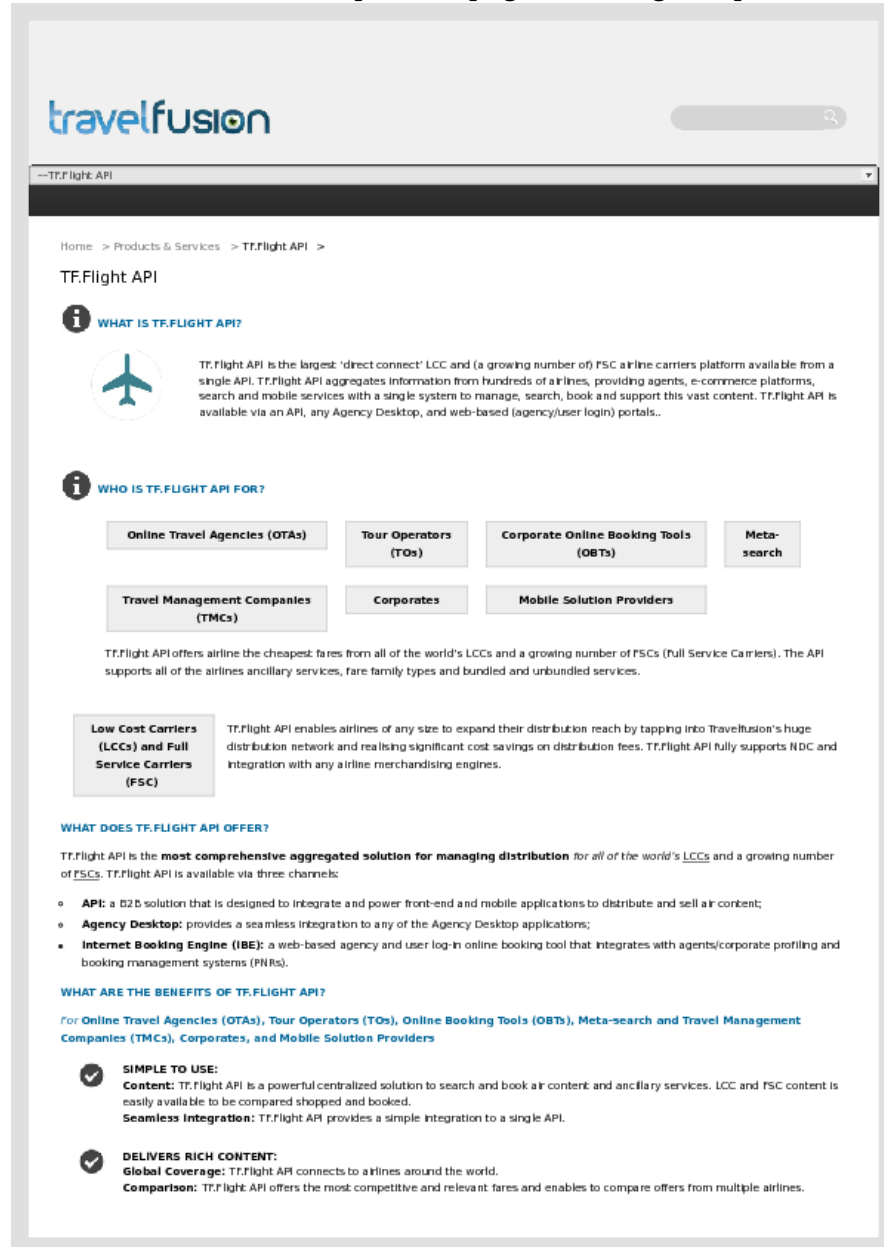
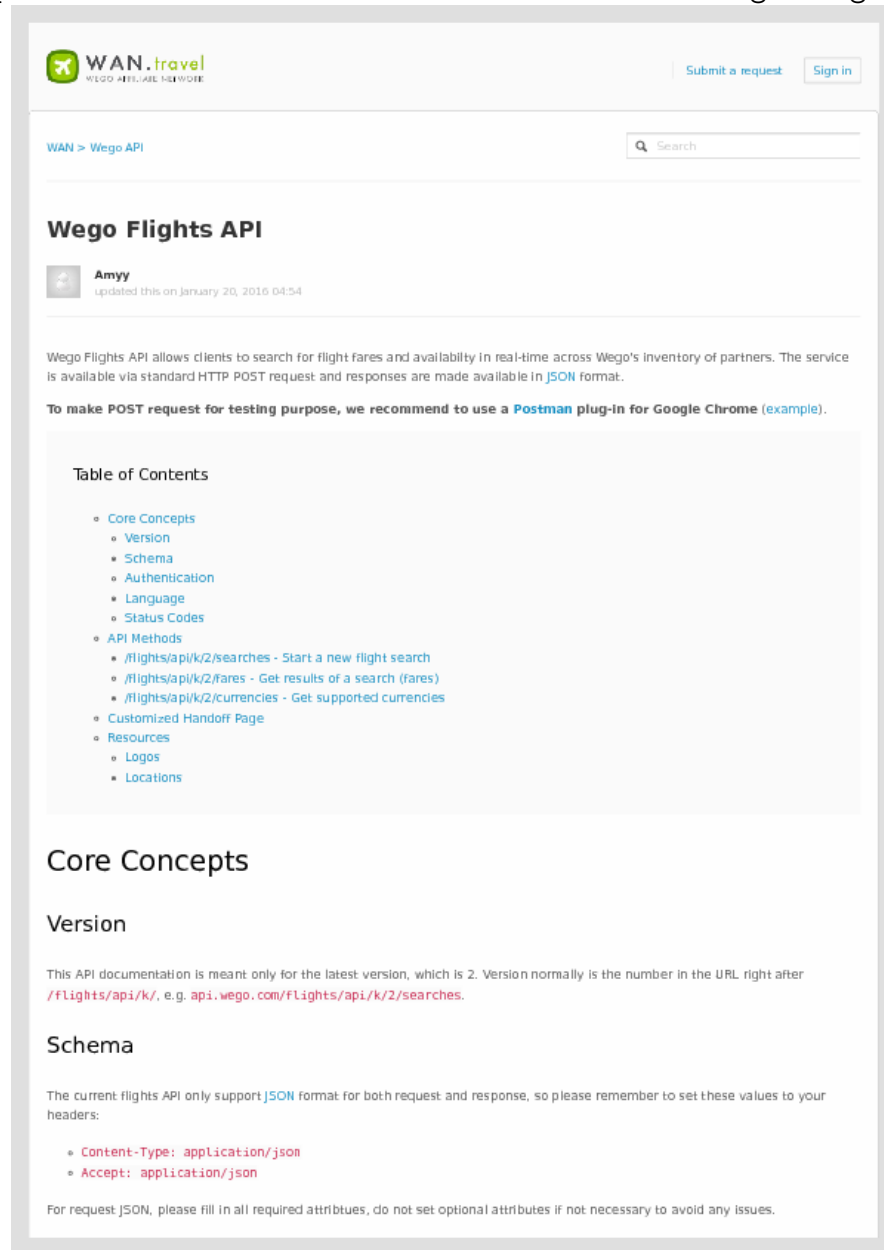


Figure 41: Travel Fusion home page.

8. Wego Flights: Wego Flights API allows clients to search for flight fares and availability in real-time across Wego's inventory of partners. The service is available via standard HTTP POST request and responses are made available in JSON format. (see Figure 42)
<http://support.wan.travel/hc/en-us/articles/200191669-Wego-Flights-API>



WAN.travel
WEGO AIRFARE NETWORK

Submit a request Sign in

WAN > Wego API

Wego Flights API

Amyy
updated this on January 20, 2016 04:54

Wego Flights API allows clients to search for flight fares and availability in real-time across Wego's inventory of partners. The service is available via standard HTTP POST request and responses are made available in [JSON](#) format.

To make POST request for testing purpose, we recommend to use a [Postman plug-in for Google Chrome](#) (example).

Table of Contents

- [Core Concepts](#)
 - [Version](#)
 - [Schema](#)
 - [Authentication](#)
 - [Language](#)
 - [Status Codes](#)
- [API Methods](#)
 - [/flights/api/k/2/searches](#) - Start a new flight search
 - [/flights/api/k/2/fares](#) - Get results of a search (fares)
 - [/flights/api/k/2/currencies](#) - Get supported currencies
- [Customized Handoff Page](#)
- [Resources](#)
 - [Logos](#)
 - [Locations](#)

Core Concepts

Version

This API documentation is meant only for the latest version, which is 2. Version normally is the number in the URL right after `/flights/api/k/`, e.g. `api.wego.com/flights/api/k/2/searches`.

Schema

The current flights API only support [JSON](#) format for both request and response, so please remember to set these values to your headers:

- `Content-Type: application/json`
- `Accept: application/json`

For request JSON, please fill in all required attributes, do not set optional attributes if not necessary to avoid any issues.

Figure 42: Wego Flights home page.

3.3 Developers

1. GeoCommons: GeoCommons is a community contributed collection of open data from around the world. Uploaded by the public, data are often from public and open government website and sources. (see Figure 43)

<http://geocommons.com/>



Figure 43: GeoCommons home page.

2. Global Data. Local Context: The best location data for Mobile Advertising, Developers, and Enterprise solutions. (see Figure 44)
<https://www.factual.com/>



Figure 44: Global Data. Local Context home page.

3. OpenStreetMap: OpenStreetMap is a map of the world, created by people and free to use under an open license. (see Figure 45)
<http://www.openstreetmap.org>

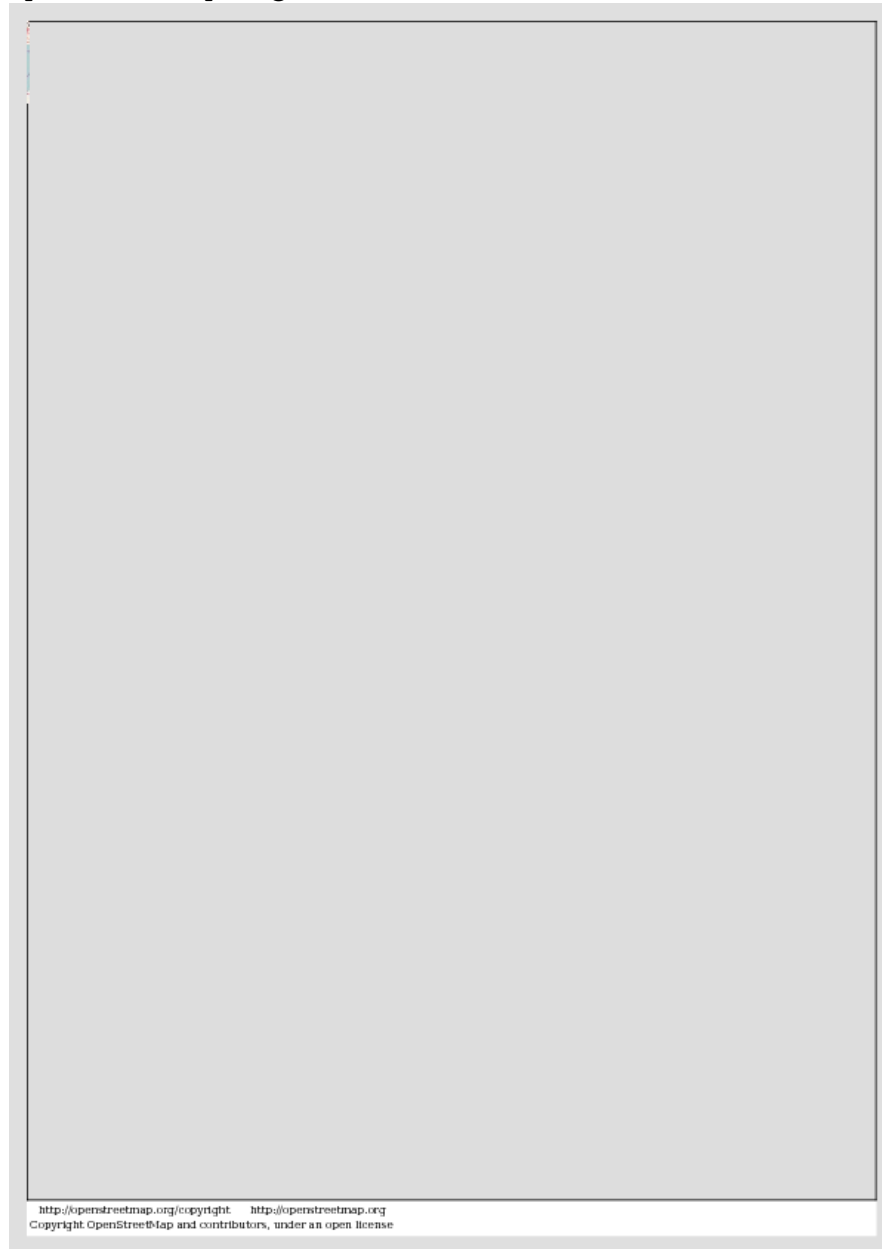


Figure 45: OpenStreetMap home page.

4. Quandl Financial and Economic Data: Quandl helps data analysts save time, effort and money by delivering high-quality financial and economic data in the precise format they need. (see Figure 46)

<https://www.quandl.com/>



Figure 46: Quandl Financial and Economic Data home page.

5. TIGER Products: TIGER = Topologically Integrated Geographic Encoding and Referencing. TIGER products are spatial extracts from the Census Bureau MAF/TIGER database, containing features such as roads, railroads, rivers, as well as legal and statistical geographic areas. The Census Bureau offers several file types and an online mapping application. (see Figure 47)

<http://www.census.gov/geo/maps-data/data/tiger.html>

The screenshot shows the 'Geography TIGER Products' page. It features a navigation bar at the top with the Census Bureau logo and links to 'Home', 'Geography', 'Maps & Data', 'TIGER Products', and 'TIGER Products'. Below the navigation bar, there is a search box and a list of product links. The main content area is titled 'Geography TIGER Products' and contains a list of product links. Below this is a table titled 'Which product should I use?' with columns for Product, Best For..., File Format, Type of Data, Level of Detail, Descriptive Attributes, and Vintages Available. The table lists several products including TIGER Line Shapefiles, TIGER Geodatabases, TIGER Line with Selected Demographic and Economic Data, Cartographic Boundary Shapefiles, KML - Cartographic Boundary Files, and TIGERweb.

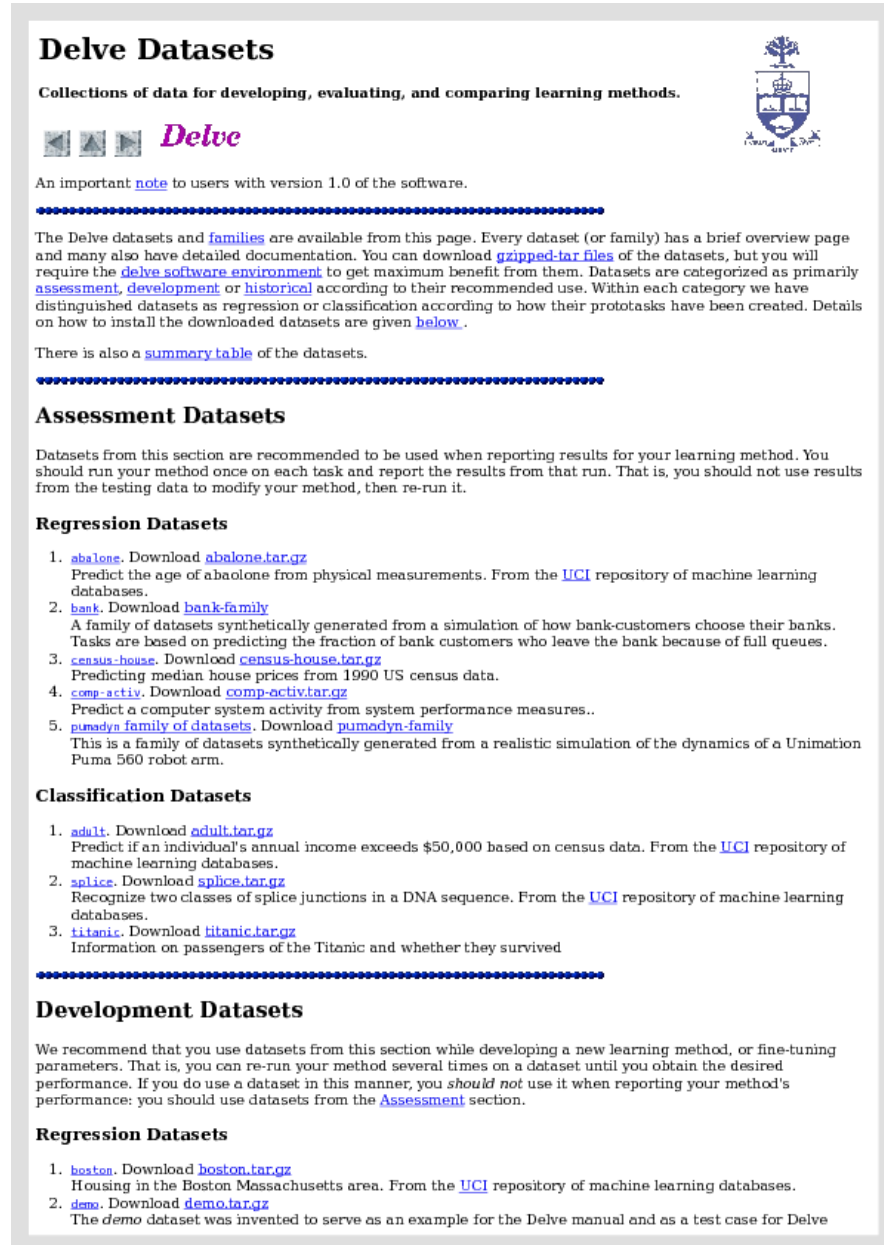
Product	Best For...	File Format	Type of Data	Level of Detail	Descriptive Attributes	Vintages Available
TIGER Line Shapefiles	Most mapping projects—this is our most comprehensive dataset . Designed for use with GIS (geographic information systems).	Shapefiles (.shp) and database files (.dbf)	Boundaries, roads, address information, water features, and more	Full detail (not generalized)	Extensive	2006 - 2015, CD 113
TIGER Geodatabases	Useful for users needing national datasets or all major boundaries for by state. Designed for use in ArcGIS. Files are extremely large.	Geodatabase (.gdb)	Boundaries, roads, address information, water features, and more	Full detail (not generalized)	Limited	2013-2015
TIGER Line with Selected Demographic and Economic Data	Data from selected attributes from the 2010 Census, 2006-2010 through 2010-2014 ACS 5-year estimates and County Business Patterns (CBP) for selected geographies. Designed for use with GIS.	Shapefiles (.shp) and Geodatabases	Boundaries, Population Counts, 2010 Housing Unit Counts, 2010 Census Demographic Profile 1 attributes, 2006-2010 through 2010-2014 ACS 5-year estimates data profiles, CBP data.	Full detail (not generalized)	Limited	2012 CBP, 2010, 2006-2010 to 2010-2014 ACS 5-Year Estimates
Cartographic Boundary Shapefiles	Small scale (limited detail) mapping projects clipped to shoreline. Designed for thematic mapping using GIS.	Shapefiles (.shp)	Selected boundaries	Less detail (generalized)	Limited	2013-2015, 2010, 2000, 1990
KML - Cartographic Boundary Files	Viewing data or creating maps using Google Earth, Google Maps, or other platforms that use KML.	KML (.kml)	Selected boundaries	Less detail (generalized)	Limited	2013-2015
TIGERweb	Viewing spatial data online or streaming to your mapping application.	Interactive viewer, HTML data files, plus REST and WMS map services	Boundaries, roads, address information, water features, and more	Detailed	Extensive	2012, 2010, 2012 ACS and 2011 ACS

Figure 47: TIGER Products home page.

3.4 Education


1. Delve Datasets: Collections of data for developing, evaluating, and comparing learning methods. (see Figure 48)

<http://www.cs.toronto.edu/~delve/data/datasets.html>



Delve Datasets

Collections of data for developing, evaluating, and comparing learning methods.

 **Delve**

An important [note](#) to users with version 1.0 of the software.

The Delve datasets and [families](#) are available from this page. Every dataset (or family) has a brief overview page and many also have detailed documentation. You can download [gzipped-tar files](#) of the datasets, but you will require the [delve software environment](#) to get maximum benefit from them. Datasets are categorized as primarily [assessment](#), [development](#) or [historical](#) according to their recommended use. Within each category we have distinguished datasets as regression or classification according to how their prototasks have been created. Details on how to install the downloaded datasets are given [below](#).

There is also a [summary table](#) of the datasets.

Assessment Datasets

Datasets from this section are recommended to be used when reporting results for your learning method. You should run your method once on each task and report the results from that run. That is, you should not use results from the testing data to modify your method, then re-run it.

Regression Datasets

1. [abalone](#). Download [abalone.tar.gz](#)
Predict the age of abalone from physical measurements. From the [UCI](#) repository of machine learning databases.
2. [bank](#). Download [bank-family](#)
A family of datasets synthetically generated from a simulation of how bank-customers choose their banks. Tasks are based on predicting the fraction of bank customers who leave the bank because of full queues.
3. [census-house](#). Download [census-house.tar.gz](#)
Predicting median house prices from 1990 US census data.
4. [comp-activ](#). Download [comp-activ.tar.gz](#)
Predict a computer system activity from system performance measures..
5. [pumadyn family of datasets](#). Download [pumadyn-family](#)
This is a family of datasets synthetically generated from a realistic simulation of the dynamics of a Unimation Puma 560 robot arm.

Classification Datasets

1. [adult](#). Download [adult.tar.gz](#)
Predict if an individual's annual income exceeds \$50,000 based on census data. From the [UCI](#) repository of machine learning databases.
2. [splice](#). Download [splice.tar.gz](#)
Recognize two classes of splice junctions in a DNA sequence. From the [UCI](#) repository of machine learning databases.
3. [titanic](#). Download [titanic.tar.gz](#)
Information on passengers of the Titanic and whether they survived

Development Datasets

We recommend that you use datasets from this section while developing a new learning method, or fine-tuning parameters. That is, you can re-run your method several times on a dataset until you obtain the desired performance. If you do use a dataset in this manner, you *should not* use it when reporting your method's performance: you should use datasets from the [Assessment](#) section.

Regression Datasets

1. [boston](#). Download [boston.tar.gz](#)
Housing in the Boston Massachusetts area. From the [UCI](#) repository of machine learning databases.
2. [demo](#). Download [demo.tar.gz](#)
The *demo* dataset was invented to serve as an example for the Delve manual and as a test case for Delve

Figure 48: Delve Datasets home page.

2. Statistics Online Computational Resource (SOCR): The Statistics Online Computational Resource (SOCR)[1] is an online multi-institutional research and education organization. SOCR designs, validates and broadly shares a suite of online tools for statistical computing, and interactive materials for hands-on learning and teaching concepts in data science, statistical analysis and probability theory. (see Figure 49)
http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data

SOCR Data

From Socr

Contents

- 1 SOCR Educational Materials - SOCR Data
- 2 Simulated data
- 3 Observed data
 - 3.1 Climate Change Data
 - 3.2 Population Data
 - 3.2 Economic, Business and Stock Market Data
 - 3.2.1 Consumer Price Index (CPI)
 - 3.2.2 Stock Market Data
 - 3.2.3 Monetary-Base Data
 - 3.2.4 Budgets and Deficits Data
 - 3.2.5 Sector Data, Population Perception Trends data
 - 3.2.6 World Peace
 - 3.4 Neuroimaging Data
 - 3.5 Biomedical Data
 - 3.6 Healthcare and Health Science Data
 - 3.7 US Census Data
 - 3.8 US Elections Data
 - 3.9 Other Data
 - 4 Multiple Interfaces to Downloading SOCR Data

SOCR Educational Materials - SOCR Data

The links below contain a number of datasets that may be used for demonstration purposes in probability and statistics education. There are two types of data - **simulated** (computer-generated using random sampling) and **observed** (research, observationally or experimentally acquired).

Simulated data

The SOCR resources provide a number of mechanisms to simulate data using computer random-number generators. Here are some of the most commonly used SOCR generators of simulated data:

- SOCR Experiments (http://socr.ucla.edu/html/SOCR_Experiments.html) - each experiment reports random outcomes, simple and population distributions and summary statistics.
- SOCR random-number generator - enables sampling of any size from any of the SOCR Distributions (<http://socr.ucla.edu/html/kbcat/>).
- SOCR Analyses (http://socr.ucla.edu/html/SOCR_Analyses.html) - all of the SOCR analyses (<http://socr.ucla.edu/html/ana>) allow random sampling from various populations appropriate for the user-specified analysis.

Observed data

The following collections include a number of real observed datasets from different disciplines, acquired using different techniques and applicable in different situations.

Climate Change Data (<http://gcmd.nasa.gov/>)

- Antarctic Ice Thickness of Mawson, Davis and Casey (01/Apr/1954 to 15/Jun/2002). Number of data points is 1630.
- Energy Resources, Production and Consumption Dataset
- California Climate Data (1980-2010)
- California and US Climate Data Snapshot

Population Data

- Human Height and Weight data
 - 25,000 Records of Adolescent Human Heights (in) and Weights (lbs)
 - Major League Baseball Player Height and Weight Data
 - Population Data by Country 2000-2100
 - Los Angeles City Neighborhood Data (from US Census)
 - Ranking of the top 100 Countries in the World based on political, economic, health, and quality-of-life factors

Economic, Business and Stock Market Data

Consumer Price Index (CPI)

- Consumer Price Index (1981-2006) - Fuel and Food Data
- Consumer Price Index (1981-2007) - One-, Two- or Three-Way ANOVA Data by Items, Months and Years
- Housing Price Index (2000-2006) (median charts)
- S&P 500 Price Index (1990-2009) (median charts)

Stock Market Data (<http://www.eoddata.com>)

- S&P 500 Systems (Ivris) Stock price (2007-2008)
- S&P 500 (2007-2008)
- US Economy by Sectors (1997-2007) and 2007-2009 Recession Data
- Ranking, Profits and Income of Fortune500 Companies (1955-2008) Dataset

Monetary-Base Data

- US Federal Reserve monetary-base data (1969-2009)
- Monthly US Economic data including monetary-base data, interest, CPI, HPI, S&P, Unemployment, Inflation, etc. (1979-2009)
- Monthly Monetary Inflation for several Countries (2002-2012)

Budgets and Deficits Data

- US Federal Budget and Deficit data (1949-2016)

Sector Data, Population Perception Trends data

- Google Web-Search Trends and Stock Market Data (2005-2011)

World Peace

- Global Peace Index Data (2001-2011)
- Wealth of Nations Data (1800-2009)

Neuroimaging Data

- Neuroimaging study of 27 Alzheimer's disease (AD) subjects, 35 normal controls (NC), and 42 mild cognitive impairment subjects (MCI)
- Alzheimer's Disease neuroimaging Data (http://www.stat.ucla.edu/~f773/box/resources_students_dbs/04/Spring/Stat223.dir/AD_Neuroimaging/Data1.html)
- Neuroimaging study of super-resolution image enhancing
- Neuroimaging study of Prefrontal Cortex Volume across Species and Tissue Types
- Normal and Schizophrenic Children's neuroimaging study
- A large Neuroimaging study of pain including visceral pain, irritable bowel syndrome, ulcerative colitis, and Crohn's disease

Figure 49: Statistics Online Computational Resource (SOCR) home page.

3.5 General

1. Bing: Bing (known previously as Live Search, Windows Live Search, and MSN Search) is a web search engine (advertised as a "decision engine") from Microsoft. (see Figure 50)
<http://www.bing.com/>

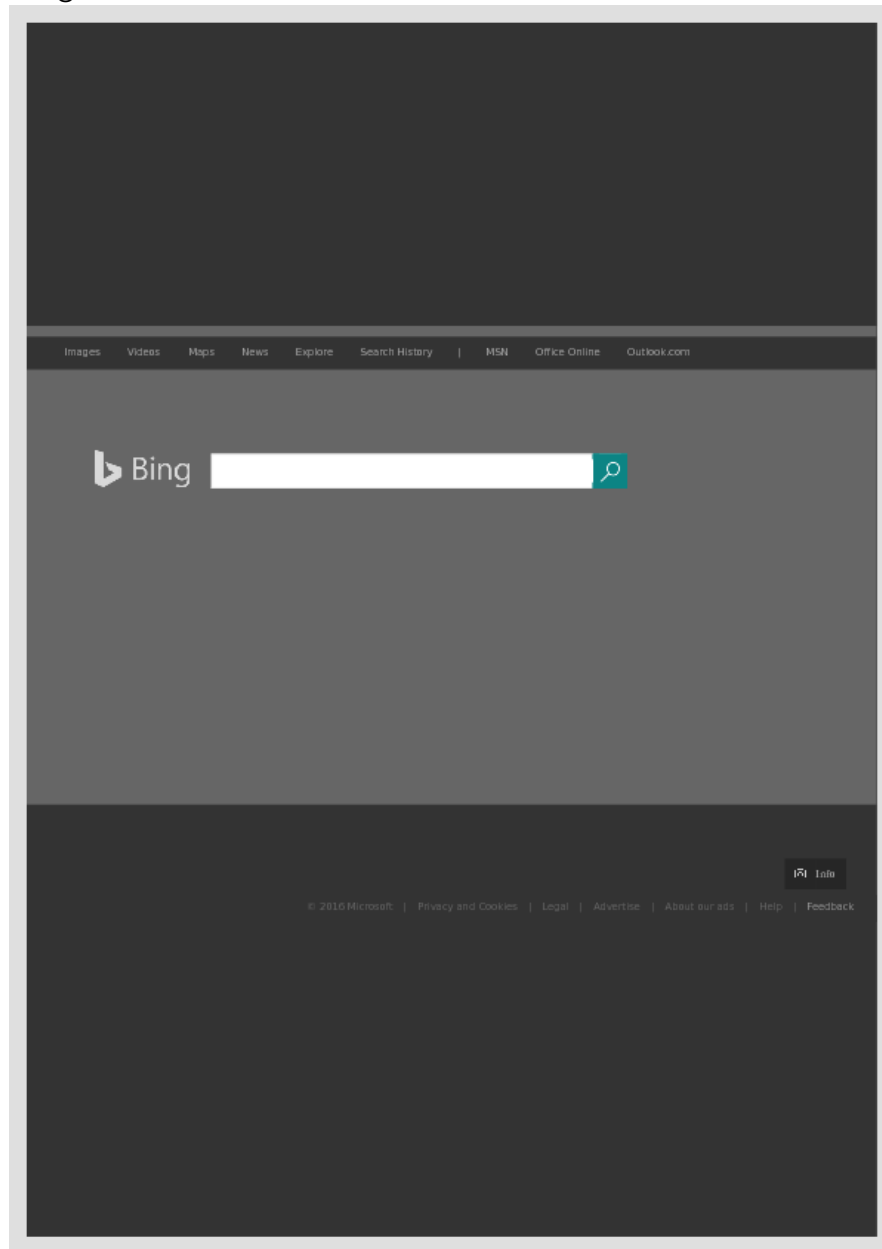


Figure 50: Bing home page.

2. Blog - AriLamstein.com: Combining Big Data and R (see Figure 51)
<http://www.arilamstein.com/blog/>

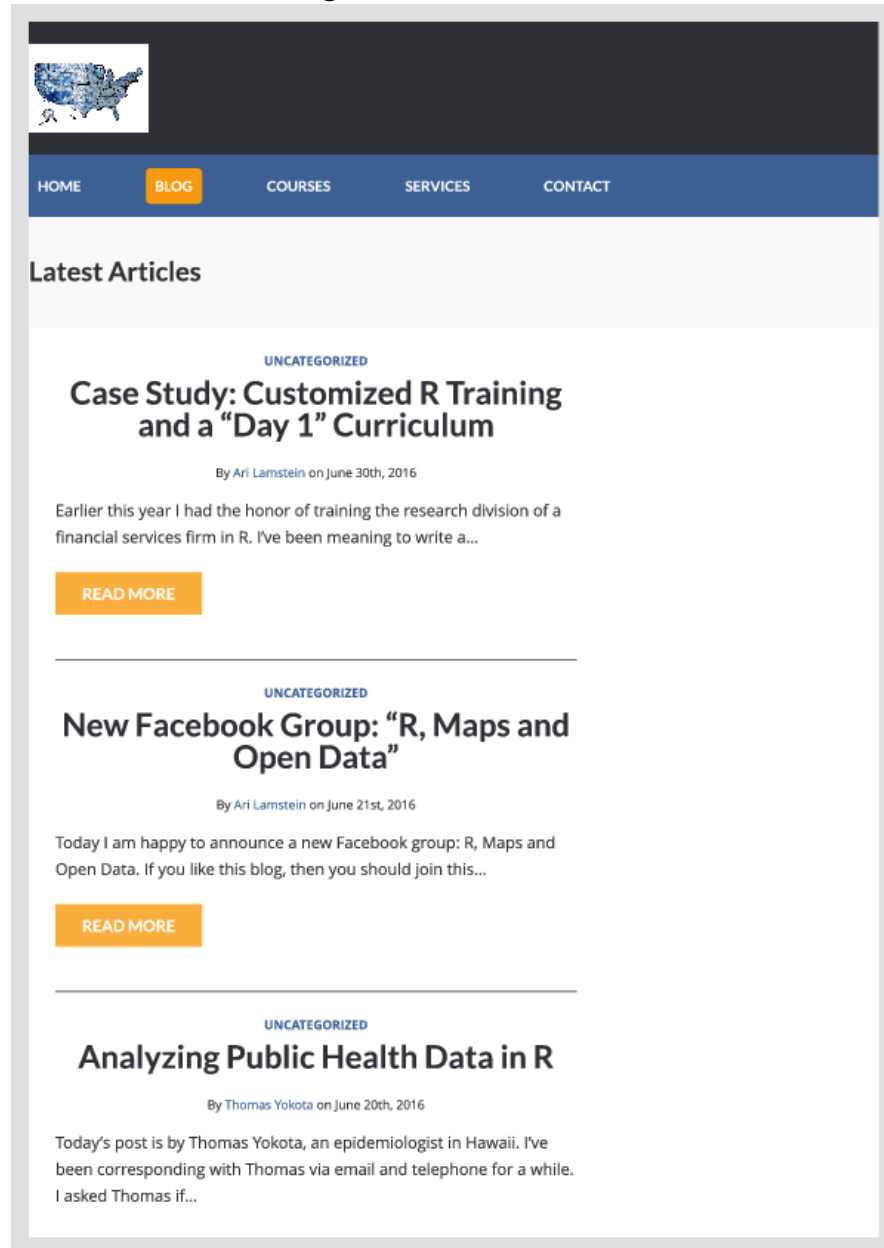


Figure 51: Blog - AriLamstein.com home page.

3. Dogpile: Dogpile is a search engine that fetches results from Google, Yahoo! and Yandex, and includes results from several other popular search engines, including those from audio and video content providers. It is a registered trademark of Blucora, Inc.. (see Figure 52) <http://www.dogpile.com/>

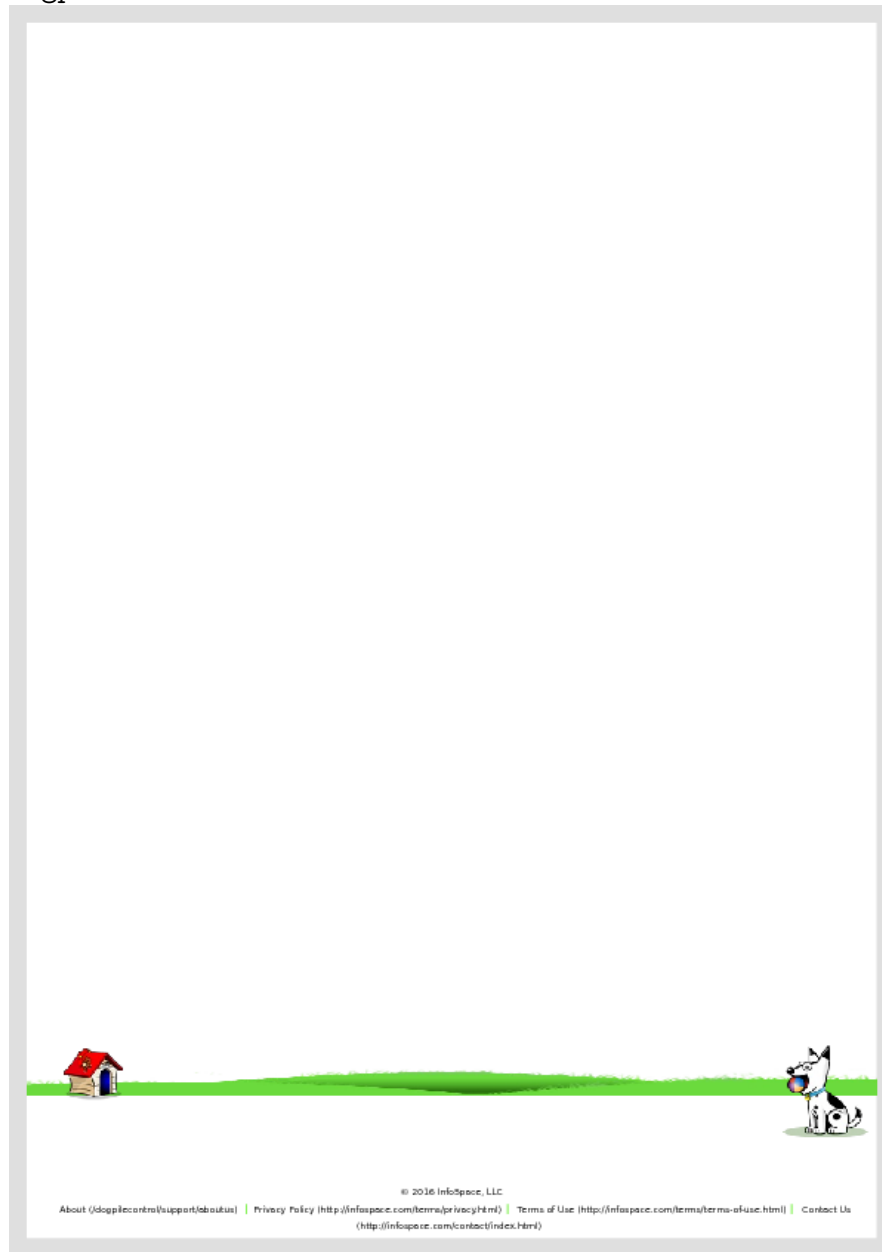


Figure 52: Dogpile home page.

4. DuckDuckGo: DuckDuckGo is an Internet search engine that emphasizes protecting searchers privacy and avoiding the filter bubble of personalized search results.[1] DuckDuckGo distinguishes itself from other search engines by not profiling its users and by deliberately showing all users the same search results for a given search term. DuckDuckGo emphasizes getting information from the best sources rather than the most sources, generating its search results from key crowd-sourced sites such as Wikipedia and from partnerships with other search engines like Yandex, Yahoo!, Bing, and Yummly. (see Figure 53)
- <https://duckduckgo.com/>

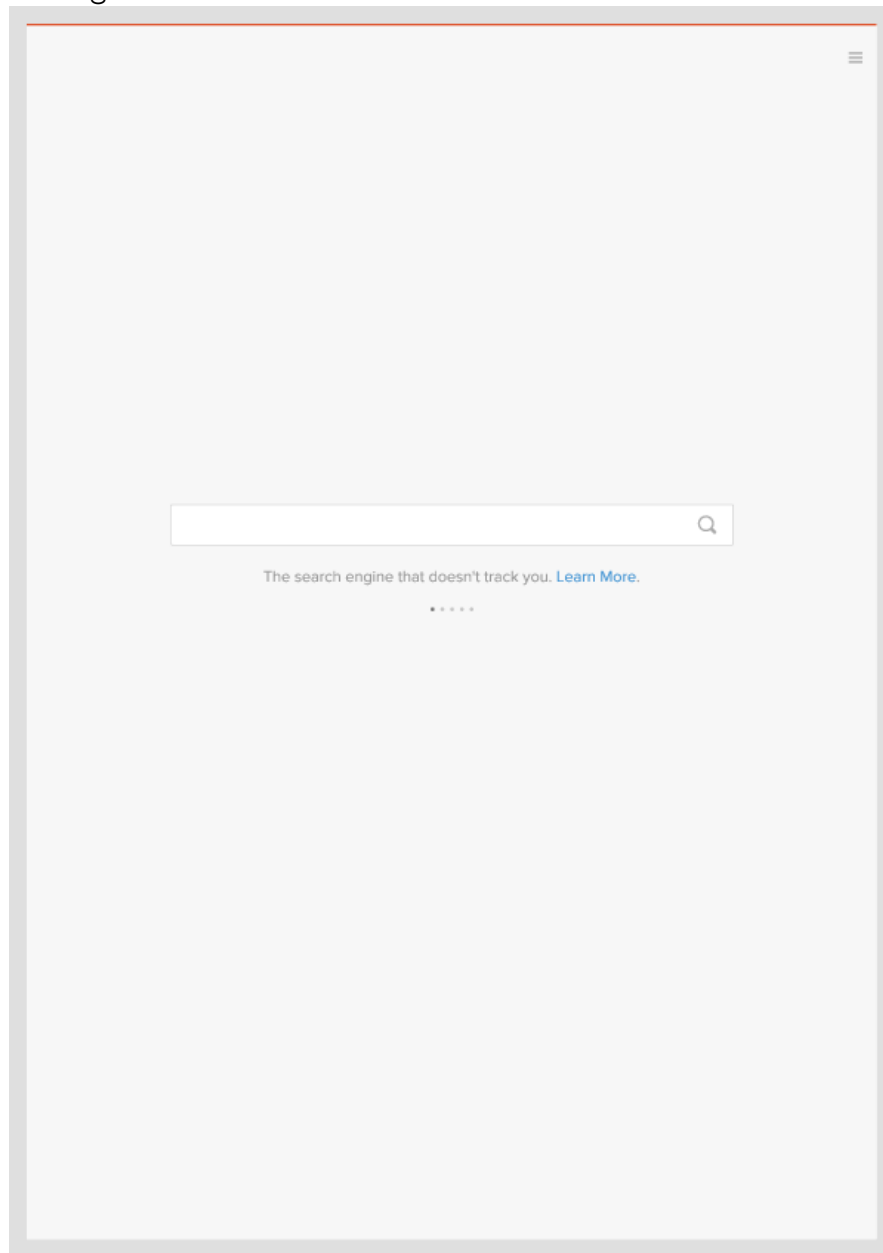


Figure 53: DuckDuckGo home page.

5. Google: Google Inc. is an American multinational technology company specializing in Internet-related services and products. These include online advertising technologies, search, cloud computing, and software. Most of its profits are derived from AdWords, an online advertising service that places advertising near the list of search results. (see Figure 54)
<http://www.google.com>

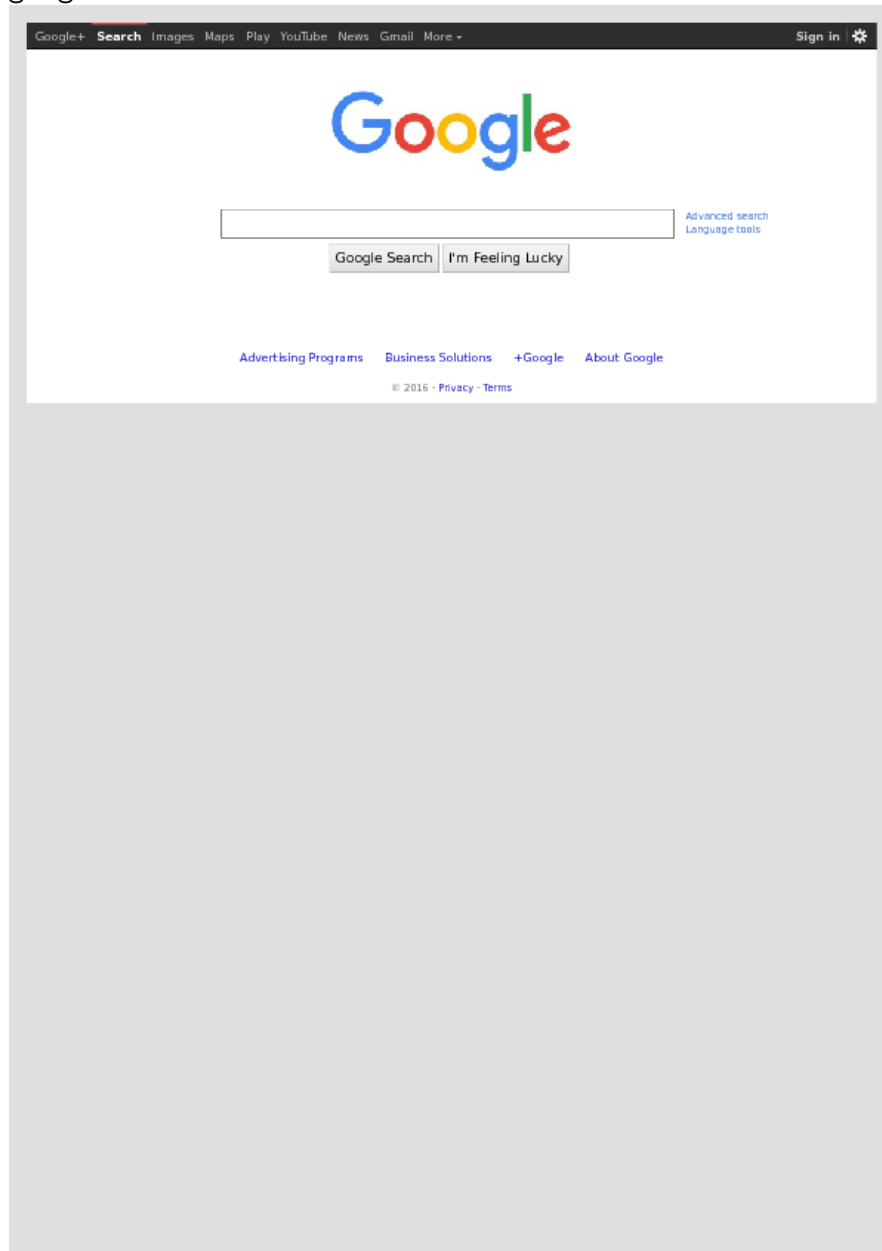


Figure 54: Google home page.

6. Our World in Data: Explore the ongoing history of human civilization at the broadest level, through research and data visualization. (see Figure 55)
<https://ourworldindata.org/>

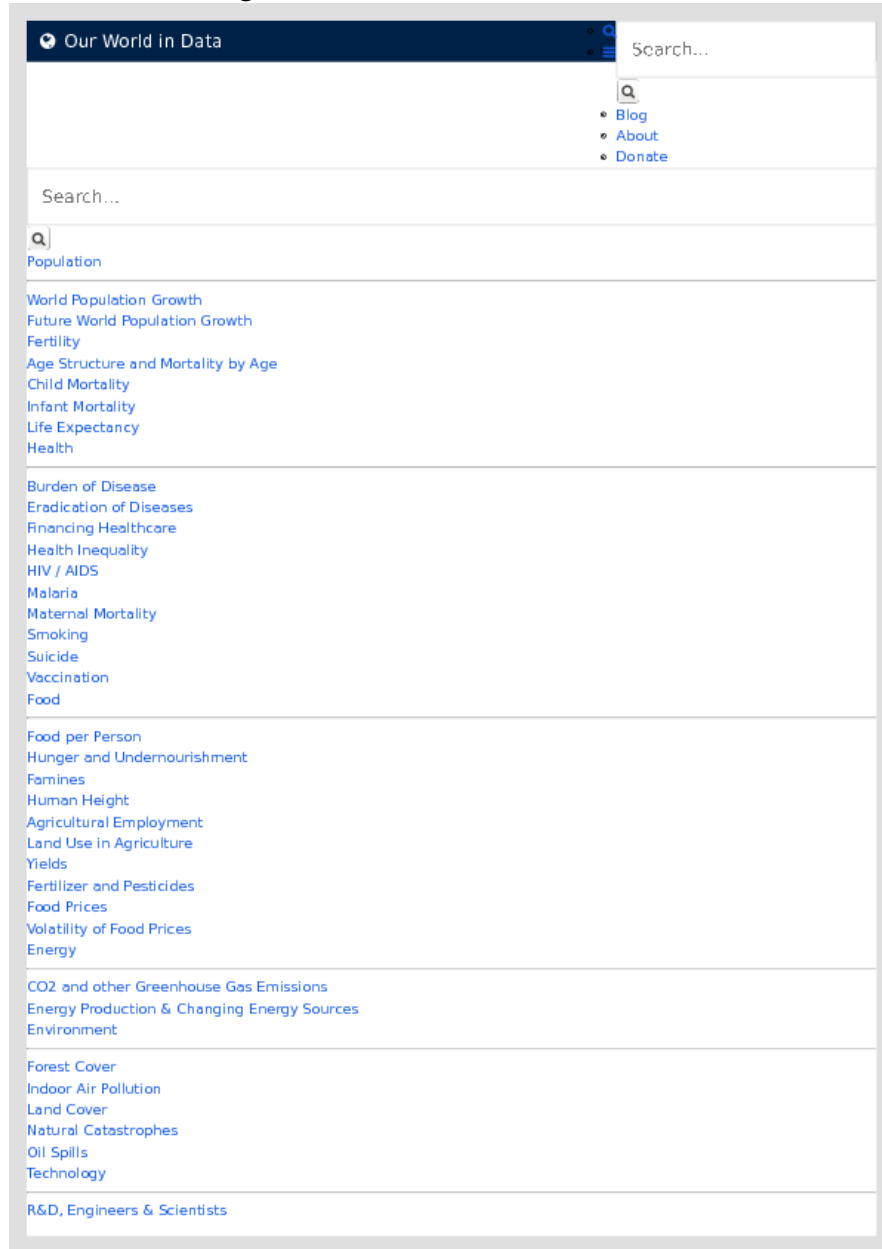


Figure 55: Our World in Data home page.

7. Stanford Large Network Dataset Collection: The SNAP library is being actively developed since 2004 and is organically growing as a result of our research pursuits in analysis of large social and information networks. Largest network we analyzed so far using the library was the Microsoft Instant Messenger network from 2006 with 240 million nodes and 1.3 billion edges. (see Figure 56)
<http://snap.stanford.edu/data/index.html>

By Jure Leskovec

STANFORD UNIVERSITY

Stanford Large Network Dataset Collection

- **Social networks** : online social networks, edges represent interactions between people
- **Networks with ground-truth communities** : ground-truth network communities in social and information networks
- **Communication networks** : email communication networks with edges representing communication
- **Citation networks** : nodes represent papers, edges represent citations
- **Collaboration networks** : nodes represent scientists, edges represent collaborations (co-authoring a paper)
- **Web graphs** : nodes represent webpages and edges are hyperlinks
- **Amazon networks** : nodes represent products and edges link commonly co-purchased products
- **Internet networks** : nodes represent computers and edges communication
- **Road networks** : nodes represent intersections and edges roads connecting the intersections
- **Autonomous systems** : graphs of the internet
- **Signed networks** : networks with positive and negative edges (friend/foe, trust/distrust)
- **Location-based online social networks** : Social networks with geographic check-ins
- **Wikipedia networks, articles, and metadata** : Talk, editing, voting, and article data from Wikipedia
- **Twitter and Memetracker** : Memetracker phrases, links and 467 million Tweets
- **Online communities** : Data from online communities such as Reddit and Flickr
- **Online reviews** : Data from online review systems such as BeerAdvocate and Amazon

SNAP networks are also available from [UP Sparse Matrix collection](#), [Visualizations of SNAP networks](#) by Tim Davis.

Open positions

We have filled all the positions for this quarter. [More info.](#)

📌 Social networks

Name	Type	Nodes	Edges	Description
ego-Facebook	Undirected	4,039	88,234	Social circles from Facebook (anonymized)
ego-Gplus	Directed	107,614	13,673,433	Social circles from Google+
ego-Twitter	Directed	81,306	1,768,149	Social circles from Twitter
soc-Epinions1	Directed	75,879	508,837	Who-trusts-whom network of Epinions.com
soc-LiveJournal1	Directed	4,847,571	68,993,773	LiveJournal online social network
soc-Pokec	Directed	1,632,803	30,622,564	Pokec online social network
soc-Slashdot0811	Directed	77,360	905,468	Slashdot social network from November 2008
soc-Slashdot0922	Directed	82,168	948,464	Slashdot social network from February 2009
wiki-Vote	Directed	7,115	103,689	Wikipedia who-votes-on-whom network
wiki-RTA	Directed, Signed	10,835	159,388	Wikipedia Requests for Adminship (with text)

📌 Networks with ground-truth communities

Name	Type	Nodes	Edges	Communities	Description
com-LiveJournal	Undirected, Communities	3,997,962	34,681,189	287,512	LiveJournal online social network
com-Friendster	Undirected, Communities	65,608,366	1,806,067,135	957,154	Friendster online social network
com-Orkut	Undirected, Communities	3,072,441	117,185,083	6,288,363	Orkut online social network
com-Youtube	Undirected, Communities	1,134,890	2,987,624	8,385	Youtube online social network
com-DBLP	Undirected, Communities	317,080	1,049,866	13,477	DBLP collaboration network
com-Amazon	Undirected, Communities	334,863	925,872	75,149	Amazon product network

📌 Communication networks

Name	Type	Nodes	Edges	Description
------	------	-------	-------	-------------

Figure 56: Stanford Large Network Dataset Collection home page.

8. UCI Machine Learning Repository: All sorts of datasets. (see Figure 57)
<https://archive.ics.uci.edu/ml/datasets.html>

The screenshot shows the UCI Machine Learning Repository website. At the top, there is a navigation bar with the UCI logo, a search bar, and links for 'About', 'Citation Policy', 'Donate a Data Set', and 'Contact'. Below the navigation bar, there is a 'Browse Through:' section with several filters:

- Default Task:** Classification (256), Regression (61), Clustering (52), Other (51)
- Attribute Type:** Categorical (137), Numerical (205), Mixed (56)
- Data Type:** Multivariate (273), Univariate (16), Sequential (35), Time-Series (62), Text (30), Domain Theory (22), Other (21)
- Area:** Life Sciences (82), Physical Sciences (43), CS / Engineering (105), Social Sciences (23), Business (19), Game (9), Other (66)
- # Attributes:** Less than 10 (84), 10 to 100 (158), Greater than 100 (56)
- # Instances:** Less than 100 (16), 100 to 1000 (128), Greater than 1000 (174)
- Format Type:** Matrix (245), Non-Matrix (105)

The main content area displays a table of 350 datasets. The table has the following columns: Name, Data Types, Default Task, Attribute Types, # Instances, # Attributes, and Year. The first few rows of the table are:

Name	Data Types	Default Task	Attribute Types	# Instances	# Attributes	Year
Abalone	Multivariate	Classification	Categorical, Integer, Real	4177	8	1995
Adult	Multivariate	Classification	Categorical, Integer	48842	14	1996
Annealing	Multivariate	Classification	Categorical, Integer, Real	798	38	
Anonymous Microsoft Web Data		Recommender-Systems	Categorical	37711	294	1998
Arrhythmia	Multivariate	Classification	Categorical, Integer, Real	452	279	1998
Artificial Characters	Multivariate	Classification	Categorical, Integer, Real	6800	7	1992
Audiology (Original)	Multivariate	Classification	Categorical	226		1987
Audiology (Standardized)	Multivariate	Classification	Categorical	226	69	1992
Auto MPG	Multivariate	Regression	Categorical, Real	398	8	1993
Automobile	Multivariate	Regression	Categorical, Integer, Real	205	26	1987
Badges	Univariate, Text	Classification		294	1	1994
Balance Scale	Multivariate	Classification	Categorical	625	4	1994
Balloons	Multivariate	Classification	Categorical	16	4	
Breast Cancer	Multivariate	Classification	Categorical	286	9	1988
Breast Cancer Wisconsin (Original)	Multivariate	Classification	Integer	699	10	1992
Breast Cancer Wisconsin (Prognostic)	Multivariate	Classification, Regression	Real	198	34	1995
Breast Cancer Wisconsin (Diagnostic)	Multivariate	Classification	Real	569	32	1995
Pittsburgh Bridges	Multivariate	Classification	Categorical, Integer	108	13	1990
Car Evaluation	Multivariate	Classification	Categorical	1728	6	1997

Figure 57: UCI Machine Learning Repository home page.

9. WolframAlpha: A specialized data and mathematical search engine. (see Figure 58)
<http://www.wolframalpha.com/>

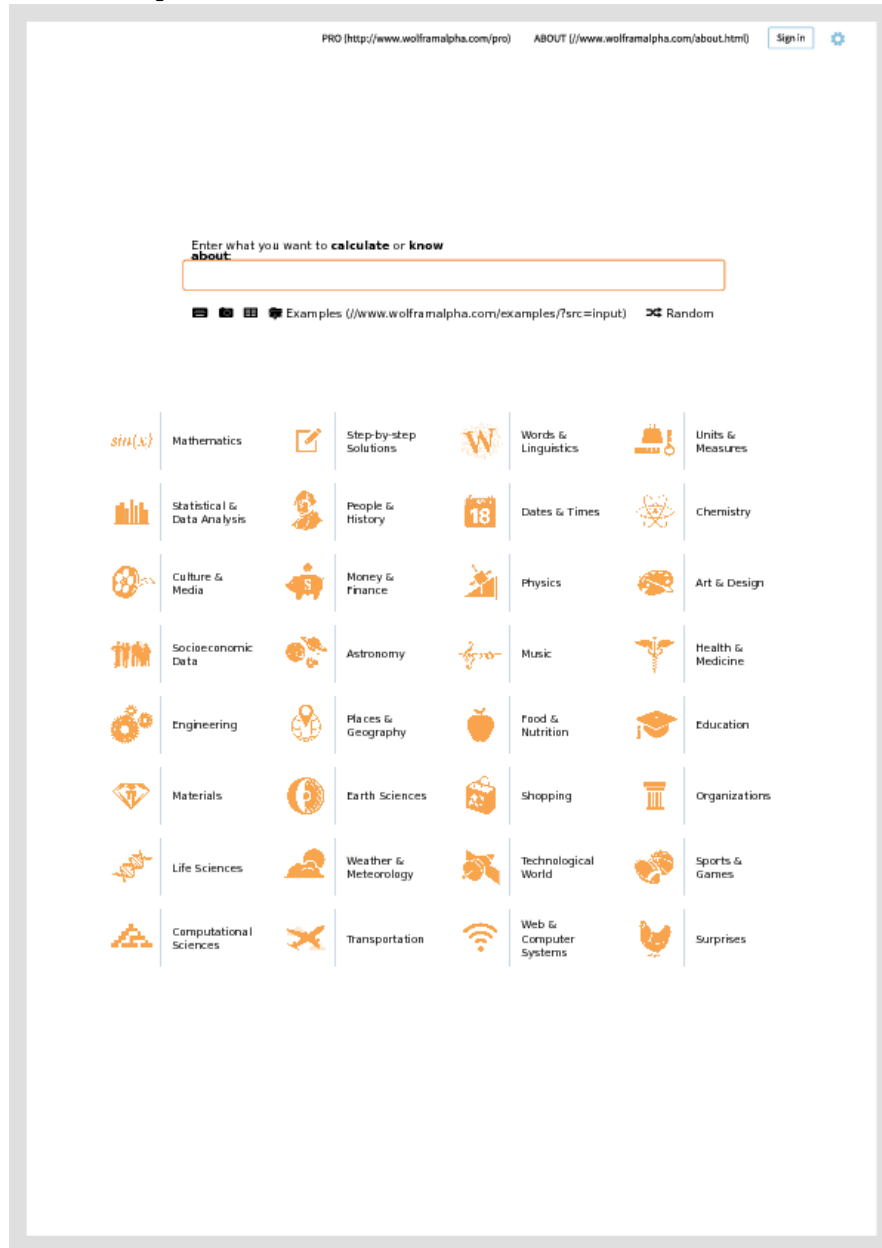


Figure 58: WolframAlpha home page.

10. Yahoo: Yahoo Inc. (styled as Yahoo!) is an American multinational technology company headquartered in Sunnyvale, California. It is globally known for its Web portal, search engine Yahoo! Search, and related services, including Yahoo! Directory, Yahoo! Mail, Yahoo! News, Yahoo! Finance, Yahoo! Groups, Yahoo! Answers, advertising, online mapping, video sharing, fantasy sports and its social media website. (see Figure 59)
<https://www.yahoo.com/>

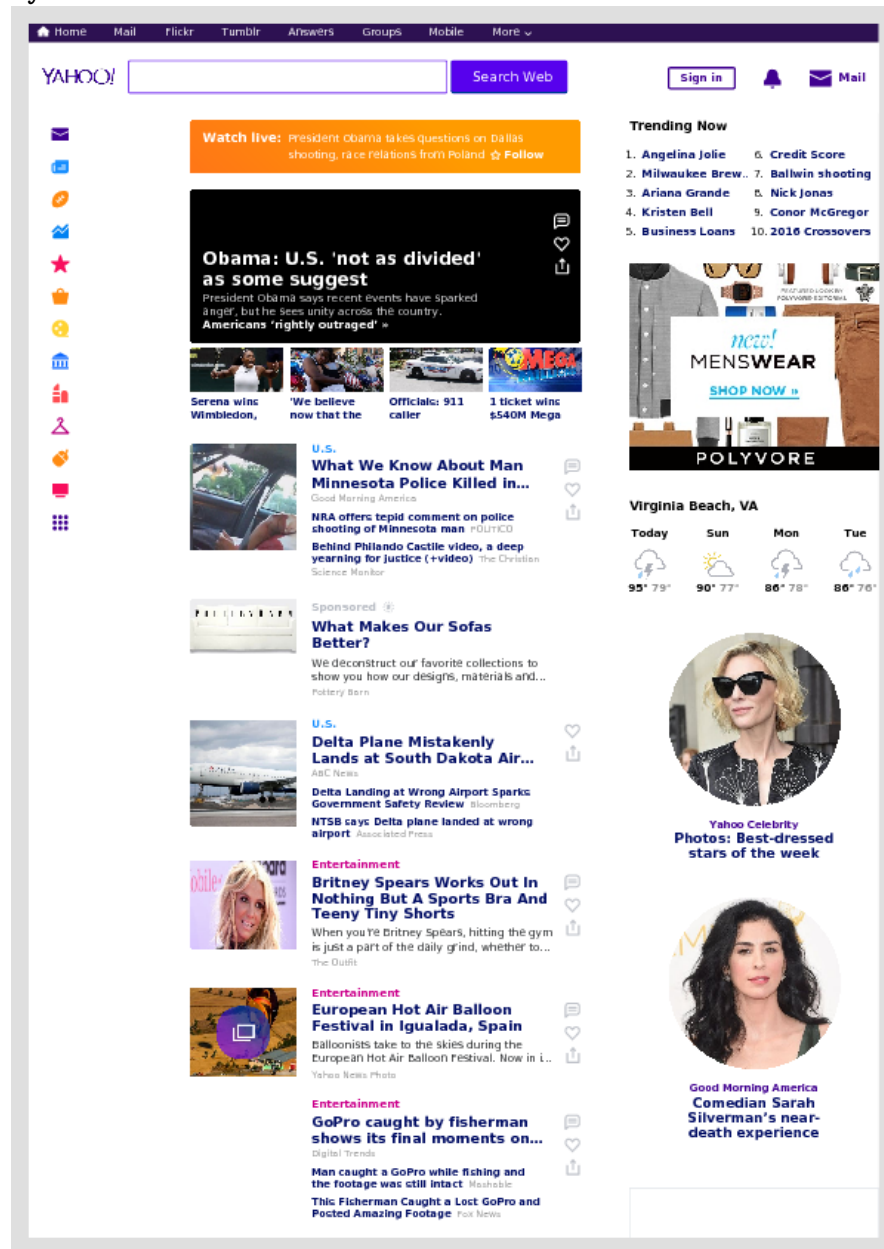


Figure 59: Yahoo home page.

3.6 Geographic information

1. GeoNames: The GeoNames geographical database covers all countries and contains over eight million place names that are available for download free of charge. (see Figure 60) <http://download.geonames.org/export/dump/>

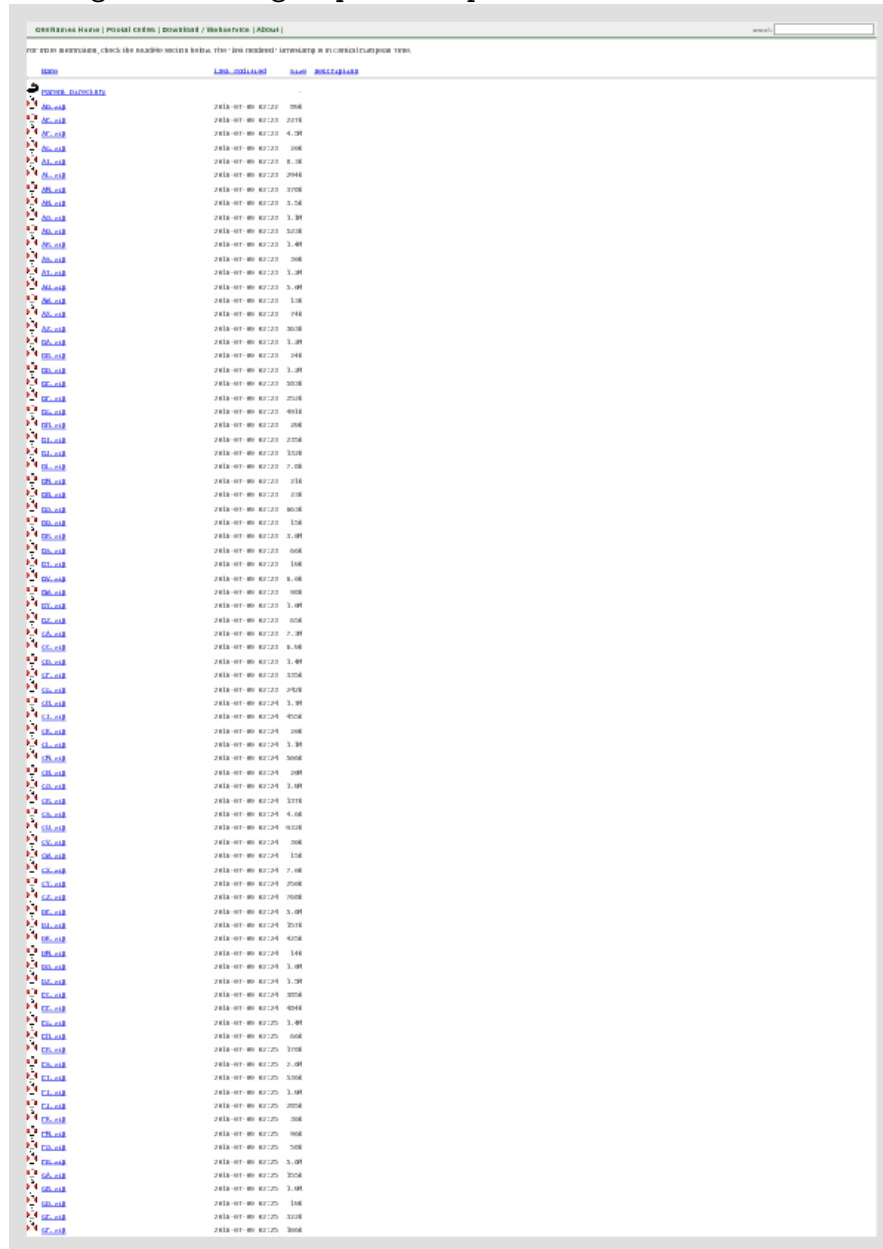


Figure 60: GeoNames home page.

2. Open Source Data and APIs — MaxMind: Add geographic information to an application using their datasets, and APIs. (see Figure 61)

<https://www.maxmind.com/en/open-source-data-and-api-for-ip-geolocation>

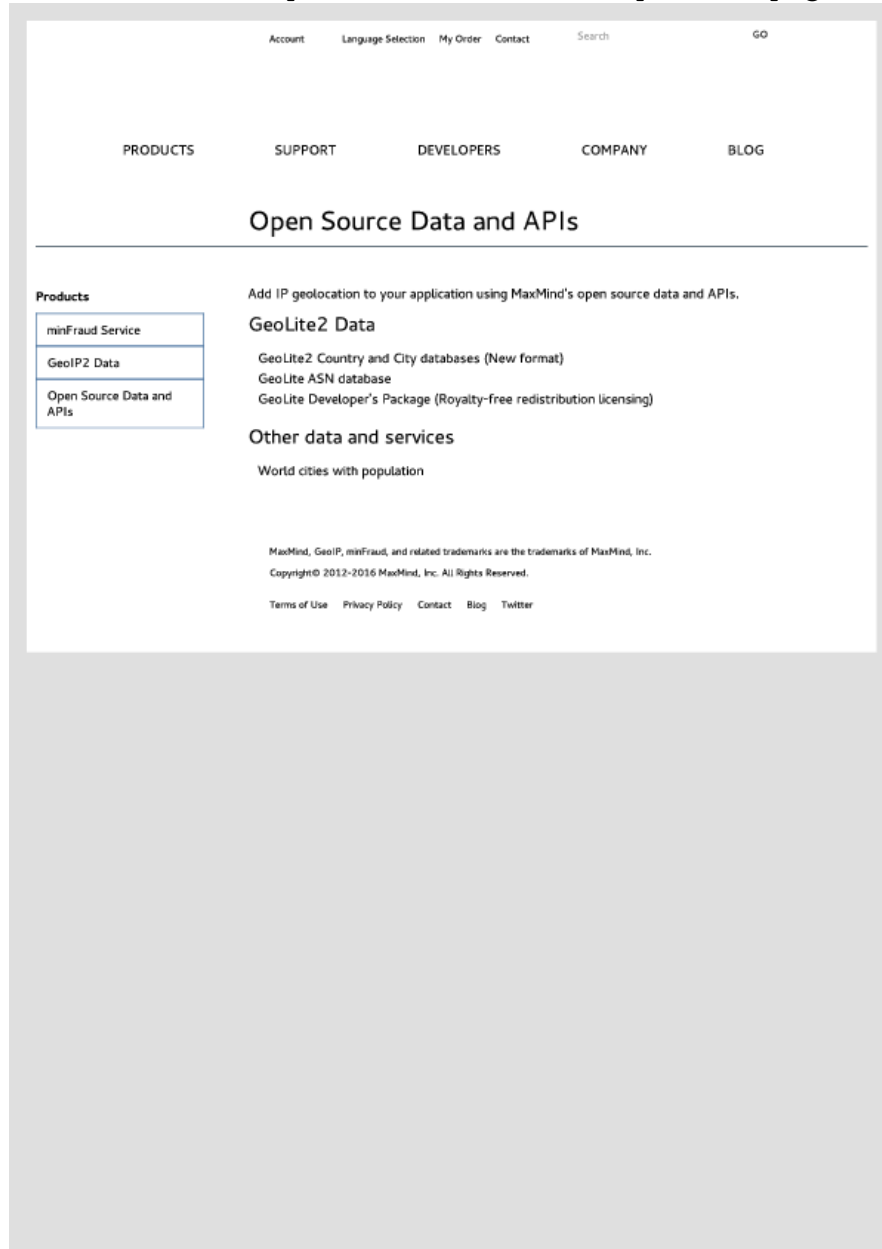


Figure 61: Open Source Data and APIs — MaxMind home page.

3. Places API - CityGrid V2 - CityGrid Media Documentation: The CityGrid Places API enables developers to create web and mobile applications that find local businesses, organizations, and points of interest by a variety of search criteria and display content associated with these places. Developers can use search results to position places on a map, perform further refinement or expansion searches, or access full details of a given place. Future versions of the Places API will allow submission of user reviews, photos, and other content. (see Figure 62)

<http://docs.citygridmedia.com/display/citygridv2/Places+API#PlacesAPI-SearchWhereNotes>

CityGrid V2 / CityGrid APIs / Content by CityGrid
Places API
Created by Librarian, last modified by Publisher on Sep 29, 2014

Introduction

The **CityGrid Places API** enables developers to create web and mobile applications that find local businesses, organizations, and points of interest by a variety of search criteria and display content associated with these places. Developers can use search results to position places on a map, perform further refinement or expansion searches, or access full details of a given place. Future versions of the Places API will allow submission of user reviews, photos, and other content.

+ If you qualify as a CityGrid partner, you can get paid for using the Places API through **Places that Pay**. Simply use the **Places that Pay** tracking features to notify CityGrid about impressions and referrals.

The Places API consists of two endpoint categories:

- **#Places Search** — Returns the set of places that match your search criteria. Only basic place information is provided with each search result.
- **#Places Detail** — Returns detailed information and associated content for a single place.

Contents

- Introduction
- Audience
- Version 2 Updates
- Places Search
 - Search Using Where
 - Where Search HTTPS Endpoint
 - Where Search Request
 - Where Search Usage Examples
 - Specifying the Where Parameter
 - Search Using Latitude and Longitude
 - LatLon Search HTTPS Endpoint
 - LatLon Search Request
 - LatLon Search Usage Examples
 - Specifying the Geography
 - Search Response
 - XML Response
 - JSON Response
 - Protocol Buffers Response
 - Spelling suggestions
 - Geocoding
 - Histograms
 - Search Error Reporting
 - Error Codes
 - XML Error Response
 - JSON Error Response
- Places Detail
 - Detail HTTPS Endpoint
 - Detail Request
 - Detail Usage Examples
 - Detail Response
 - XML Response
 - JSON Response
 - Protocol Buffers Response
 - Impression Url
 - Detail Error Reporting
 - Error Codes
 - XML Error Response
 - JSON Error Response

Audience

The Places API is intended for developers of web and mobile applications who want to give their applications the ability to search for and display profiles of local businesses and points of interest.

Version 2 Updates

> The following are changes for version 2, updated from version 1

- Added support for SQL
- New endpoint URLs beginning with <http://api.citygridmedia.com/content/places/v2/>.
- New parameters for search requests:
 - histograms
 - lat2
 - lon2
 - has_offers
 - i (impression_id)
- Changes to the Places Search response:
 - Added `impression_id`.
 - If there is no `featured`, an empty element is still provided in the response.
 - If there is no `distance`, an empty element is still provided in the response.

Figure 62: Places API - CityGrid V2 - CityGrid Media Documentation home page.

3.7 Geographic location

1. MaxMind Open Source Data and APIs: Add IP geolocation to your application using MaxMind's open source data and APIs. (see Figure 63)
<https://www.maxmind.com/en/open-source-data-and-api-for-ip-geolocation>

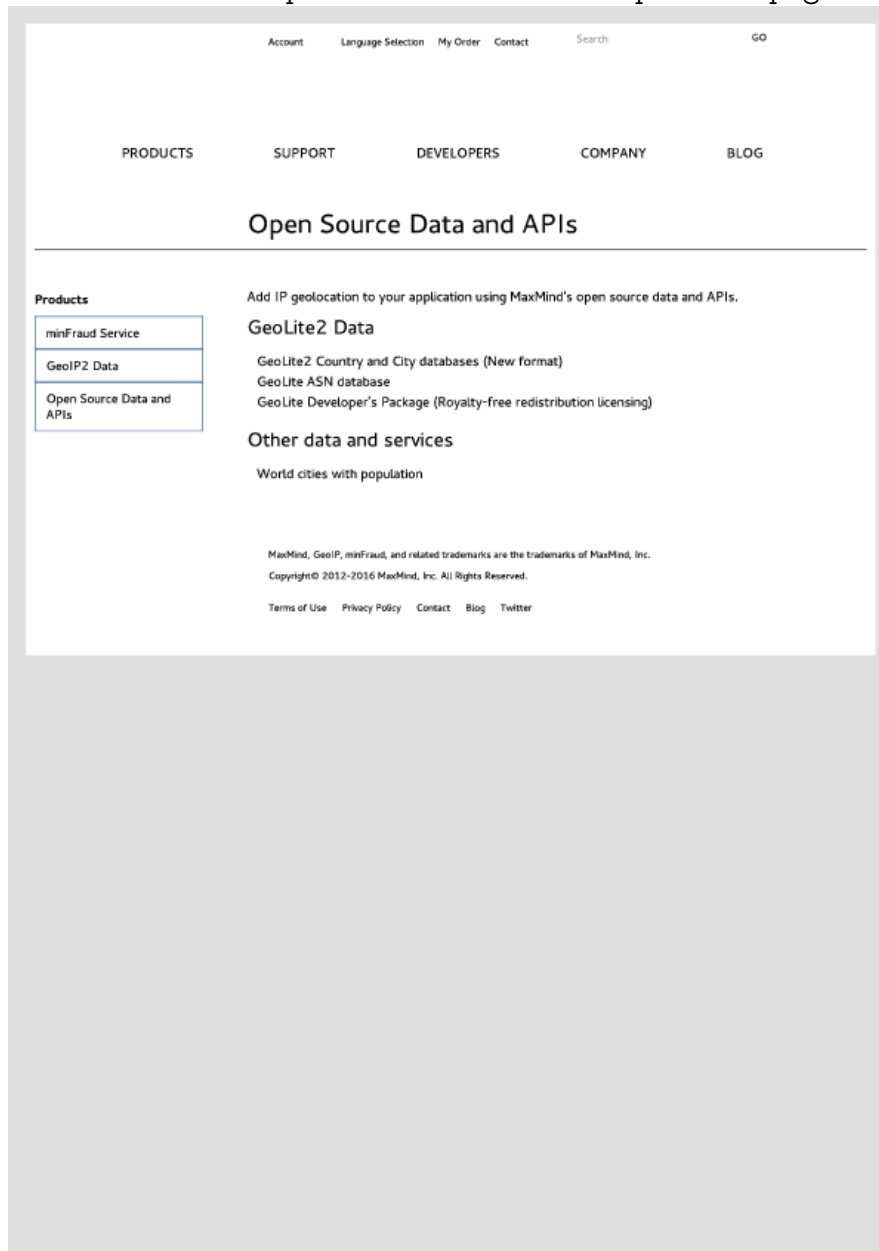


Figure 63: MaxMind Open Source Data and APIs home page.

3.8 Government

1. Austin Texas Data Portal: The Data Portal is your open government resource. Data portal benefits include: transparency, user-friendly data presentation and opportunities for community app development. (see Figure 64)
<https://data.austintexas.gov/>

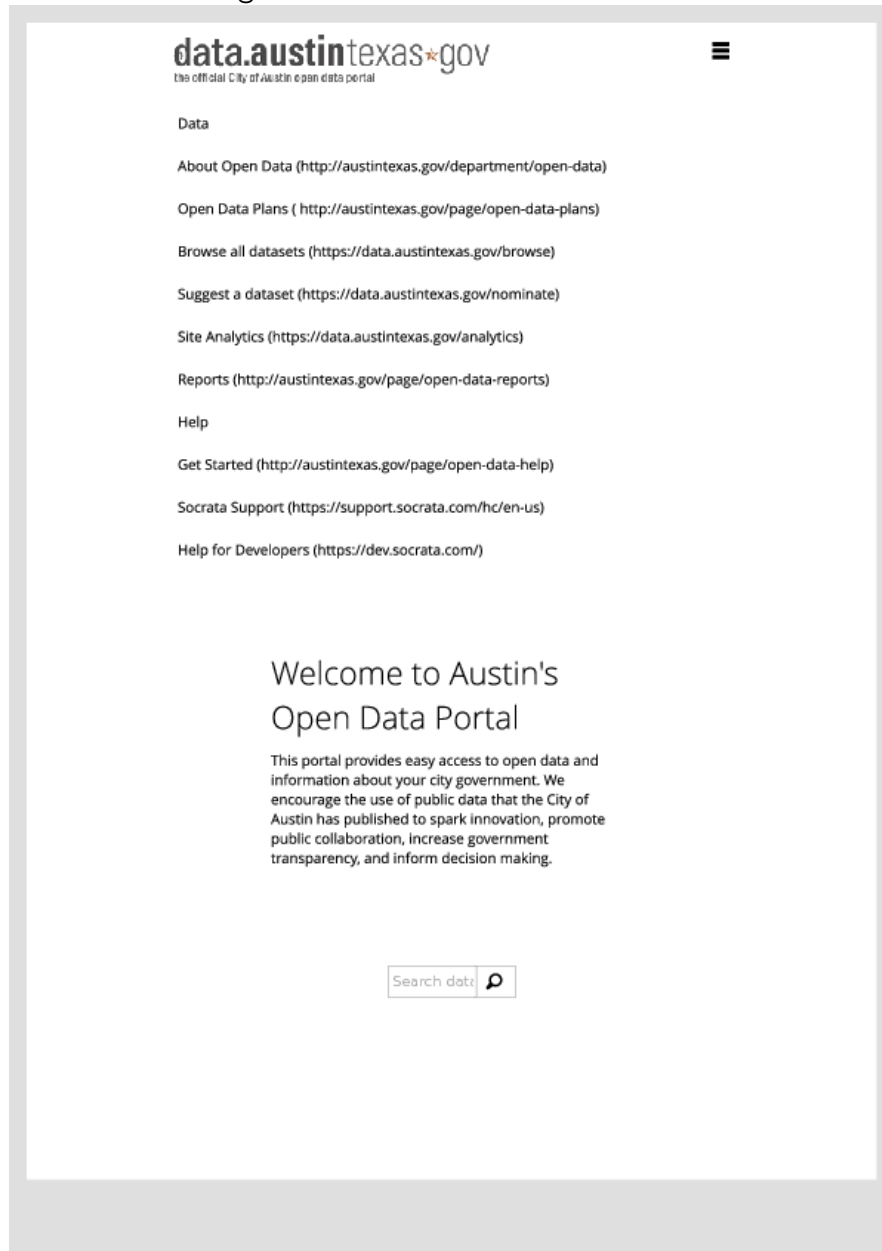


Figure 64: Austin Texas Data Portal home page.

2. Bureau of Economic Analysis: The Bureau of Economic Analysis (BEA) promotes a better understanding of the U.S. economy by providing the most timely, relevant, and accurate economic accounts data in an objective and cost-effective manner. (see Figure 65)

<http://www.bea.gov/index.htm>

(/index.htm) Go

(<http://www.commerce.gov>)

[Home \(/index.htm\)](#)
[National \(/national/index.htm\)](#)
[International \(/international/index.htm\)](#)

[Regional \(/regional/index.htm\)](#)
[Industry \(/industry/index.htm\)](#)
[Interactive Data \(/itable/\)](#)

[Contact Us \(/contacts/search.htm\)](#)
[FAQs \(/faq/index.cfm\)](#)
[About BEA \(/about/index.htm\)](#)

7/7/2016

Latest Release

Real Personal Income for States and Metropolitan Areas

(/newsreleases/regional/rpp/rpp_newsrelease.htm)

6/28/2016

Real GDP
+1.1% in Q1 2016

(</newsreleases/national/gdp/gdpnewsrelease.htm>)

6/29/2016

Personal Income
+0.2% in May 2016

(</newsreleases/national/pi/pinewsrelease.htm>)

7/6/2016

Int'l Trade in Goods and Services

Deficit increased to \$41.1 billion in May 2016(p) from \$37.4 billion in April 2016(r).

(</newsreleases/international/trade/tradnewsrelease.htm>)

6/16/2016

U.S. Int'l Transactions

Current-account deficit increased \$11.3 billion to \$124.7 billion in Q1 2016(p).

(</newsreleases/international/transactions/transnewsrelease.htm>)

2016 News Release Schedule (</newsreleases/2016rd.htm>)

Video

Looking for Videos on BEA Economic Indicators? (<https://www.youtube.com/watch?v=6SN6rPDf7og>)

Real Disposable Personal Income and Real Consumer Spending (<https://www.youtube.com/watch?v=zhMBkpL00kc>)

New International FDI Data Now Available (<https://www.youtube.com/watch?v=zVqBHQChLM>)

New International Data Tool (<https://www.youtube.com/watch?v=xgLdKJV-g2g>)

BEA Blog

Jun 29, 2016
Why Do Old GDP Numbers Keep Changing?
(<https://blog.bea.gov/2016/06/29/why-do-old-gdp-numbers-keep-changing/>)

Jul 07, 2016
Real Personal Income for States, 2014
(<https://blog.bea.gov/2016/07/07/real-personal-income-for-states-2014/>)

Jul 06, 2016
May 2016 Trade Gap is \$41.1 Billion

Figure 65: Bureau of Economic Analysis home page.

3. Bureau of Labor Statistics: The Bureau of Labor Statistics of the U.S. Department of Labor is the principal Federal agency responsible for measuring labor market activity, working conditions, and price changes in the economy. Its mission is to collect, analyze, and disseminate essential economic information to support public and private decision-making. As an independent statistical agency, BLS serves its diverse user communities by providing products and services that are objective, timely, accurate, and relevant. (see Figure 66) <http://www.bls.gov/>

The screenshot shows the Bureau of Labor Statistics (BLS) home page. At the top, there is a red navigation bar with the BLS logo and various links. Below this, the main content area is divided into several sections:

- MONTHLY LABOR REVIEW:** A featured article titled "Labor productivity growth in elementary and secondary school services" with a photo of children in a classroom.
- LATEST NUMBERS:** A list of key economic indicators such as Consumer Price Index (CPI), Unemployment Rate, Payroll Employment, Average Hourly Earnings, and more.
- REGIONAL HOMEPAGES:** A section for browsing data by region, with a list of regional names like New England, New York, New Jersey, etc.
- THE ECONOMICS DAILY:** A list of recent news items related to employment and the economy.
- CAREER INFORMATION:** Links to resources like the Occupational Outlook Handbook and Career Outlook.
- PUBLICATIONS:** Links to the Monthly Labor Review and other BLS publications.
- RESOURCES FOR...:** A section for students and teachers, including links to the Student and Teacher resources.
- DID YOU KNOW?:** A fact box stating that many operations research analysts who work with the military are veterans of the U.S. Armed Forces.

Figure 66: Bureau of Labor Statistics home page.

4. Census Bureau Economic Statistics: Historical and current information about various industries. (see Figure 67)
<https://www.census.gov/econ/>

The screenshot shows the Census Bureau Economic Statistics website. At the top, there is a navigation bar with links to the U.S. Department of Commerce, Blog, Index A-Z, and other resources. Below this is a search bar and a breadcrumb trail: "You are here: Census.gov (/) > Business & Industry (/econ/isp/)".

The main content area is titled "Census Bureau Economic Statistics" and is divided into three columns:

- The Economic Census (/econ/census/index.html)**: Provides data for every five years (2002, 2007, 2012, etc.) for every industry, with statistics for U.S., states, metro areas, counties, and cities. It includes links for the 2012 Economic Census and search databases for 2007 and 2002.
- Economic Indicators (https://www.census.gov/economic-indicators/)**: Offers monthly and quarterly data for selected sectors, along with national statistics. It includes a calendar view of indicator releases and links to current data and time-series charts.
- Other Economic Programs (other_econ.html)**: Lists various services such as Annual & Quarterly Services, County Business Patterns, Enterprise Statistics, and Monthly & Annual Wholesale Trade.

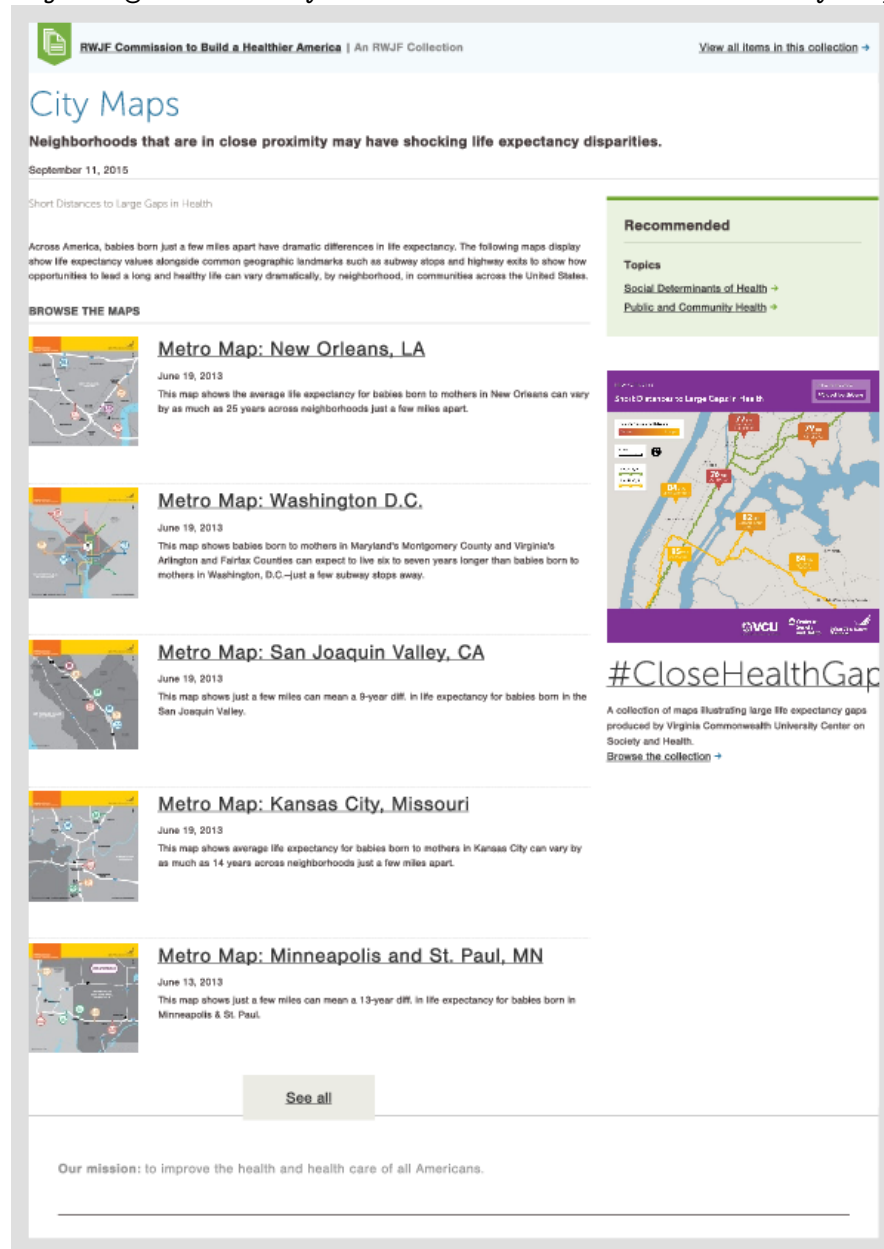
A "Latest Economic Indicator" table is prominently displayed:

Latest Economic Indicator	Previous	Current
New Residential Construction (starts) (indicator/www/mq/c20_cum.gif) Privately-owned housing starts in March 2014 were at a seasonally adjusted annual rate of 946,000. This is 2.8 percent (+/- 14.7%) above the revised February 2014 estimate of 920,000.	+1.9* February 2014 (r) % change	+2.8* March 2014 % change

At the bottom of the page, there is a grid of links for various sections: ABOUT US, FIND DATA, BUSINESS & INDUSTRY, PEOPLE & HOUSEHOLDS, SPECIAL TOPICS, and NEWSROOM. A footer contains the source information: "Source: U.S. Census Bureau | Business & Industry | 301.763-2547 | econ@census.gov | Last Revised: August 07, 2015".

Figure 67: Census Bureau Economic Statistics home page.

5. City Maps - Robert Wood Johnson Foundation: A discussion about how your life expectancy can be estimated by your subway stop. (see Figure 68)
<http://www.rwjf.org/en/library/articles-and-news/2015/09/city-maps.html>



RWJF Commission to Build a Healthier America | An RWJF Collection [View all items in this collection →](#)

City Maps

Neighborhoods that are in close proximity may have shocking life expectancy disparities.

September 11, 2015

Short Distances to Large Gaps in Health

Across America, babies born just a few miles apart have dramatic differences in life expectancy. The following maps display show life expectancy values alongside common geographic landmarks such as subway stops and highway exits to show how opportunities to lead a long and healthy life can vary dramatically, by neighborhood, in communities across the United States.

BROWSE THE MAPS

Metro Map: New Orleans, LA
June 19, 2013
This map shows the average life expectancy for babies born to mothers in New Orleans can vary by as much as 25 years across neighborhoods just a few miles apart.

Metro Map: Washington D.C.
June 19, 2013
This map shows babies born to mothers in Maryland's Montgomery County and Virginia's Arlington and Fairfax Counties can expect to live six to seven years longer than babies born to mothers in Washington, D.C.—just a few subway stops away.

Metro Map: San Joaquin Valley, CA
June 19, 2013
This map shows just a few miles can mean a 9-year diff. in life expectancy for babies born in the San Joaquin Valley.

Metro Map: Kansas City, Missouri
June 19, 2013
This map shows average life expectancy for babies born to mothers in Kansas City can vary by as much as 14 years across neighborhoods just a few miles apart.

Metro Map: Minneapolis and St. Paul, MN
June 13, 2013
This map shows just a few miles can mean a 13-year diff. in life expectancy for babies born in Minneapolis & St. Paul.

[See all](#)

Recommended

Topics

- [Social Determinants of Health →](#)
- [Public and Community Health →](#)

#CloseHealthGap

A collection of maps illustrating large life expectancy gaps produced by Virginia Commonwealth University Center on Society and Health.
[Browse the collection →](#)

Our mission: to improve the health and health care of all Americans.

Figure 68: City Maps - Robert Wood Johnson Foundation home page.

6. Datasets - data.sa.gov.au: A collection of datasets made available by the South Australia government. (see Figure 69)
<https://data.sa.gov.au/data/dataset>

The screenshot shows the Data SA website interface. At the top, there are navigation links for 'Log in (/data/user/login)', 'Register (/data/user/register)', and 'Statistics (/data/stats)'. Below this is a blue header with the 'Data SA' logo and the text 'South Australia Government Data Directory'. A search bar is located in the top right of the header. The main content area is titled 'Datasets (/data/dataset)' and features a search bar with the placeholder text 'Search datasets...'. Below the search bar, there is a dropdown menu for 'Order by' set to 'Last Modified'. A large heading indicates '1,065 datasets found', followed by a 'Filter Results' button. The page lists several datasets, each with a title, a brief description, and available file formats. The datasets listed are: 'School Zones for South Australian Government High Schools (/data/dataset/school-zones-for-south-australian-government-high-schools-2015)', 'Water Catchments Map (/data/dataset/catchments-in-the-city-of-onkaparinga-map)', 'Accounts Payable - Number of Business Days for Scan To Workflow Process FY15 (/data/dataset/procapp29)', 'Drillholes within South Australia (/data/dataset/drillholes-within-south-australia)', 'Accounts Payable - Total Invoices Paid Via Basware FY15 (/data/dataset/procapp26)', and 'Accounts Payable - Number of Invoices Escalated FY15 (/data/dataset/procapp33)'. Each dataset entry includes a description and links for SHP, KMZ, GeoJSON, metadata, and HTML formats.

Figure 69: Datasets - data.sa.gov.au home page.

7. Earthquake Hazards Program: The Earthquake Notification Service (ENS) is a free service that sends you automated notifications to your email or cell phone when earthquakes happen. (see Figure 70)

<http://earthquake.usgs.gov/earthquakes/feed/v1.0/>

The screenshot shows the 'Feeds & Notifications' page of the U.S. Geological Survey - Earthquake Hazards Program. The page is organized into several sections:

- Real-time Feeds**
 - ATOM Syndication**: A basic syndication format supported by a variety of feed readers. This is a good option for casually subscribing to earthquake information.
 - Google Earth™ KML**: This feed format is suitable for loading into applications that understand Keyhole Markup Language (KML) such as Google Earth™.
 - Spreadsheet Format**: A simple text format suitable for loading data into spreadsheet applications like Microsoft Excel™. This is a good option for manual scientific analysis.
 - QuakeML**: A flexible, extensible and modular XML representation of seismological data which is intended to cover a broad range of fields of application in modern seismology.
- Real-time Notifications**
 - Earthquake Notification Service**: The Earthquake Notification Service (ENS) is a free service that sends you automated notifications to your email or cell phone when earthquakes happen.
 - Tweet Earthquake Dispatch**: Tweet Earthquake Dispatch (TED) offers two Twitter accounts. On average, each account will produce about one tweet per day.
- For Developers**
 - [API Documentation - EQ Catalog](#)
 - [GeoJSON Summary Feed](#)
 - [GeoJSON Detail Feed](#)
 - [Change Log](#)
 - [Feed Lifecycle Policy](#)
 - [Developers Corner](#)
 - [Glossary - Earthquake Catalog Data Terms](#)
 - [Mailing List - Announcements](#)
 - [Mailing List - Forum/Questions](#)

Figure 70: Earthquake Hazards Program home page.

8. FAA Web Services, Airport Service: Serves airport status and delay information from the Air Traffic Control System Command Center (ATCSCC) as displayed on <http://fly.faa.gov/> (see Figure 71)

<http://services.faa.gov/docs/services/airport/>

Federal Aviation Administration

Airport Service

Description
Serves airport status and delay information from the Air Traffic Control System Command Center (ATCSCC) as displayed on <http://fly.faa.gov/> (<http://fly.faa.gov/>).

Methods
Airport Status
Gets the airport status for any major airport, including known delays and weather data from NOAA.

URI:
<http://services.faa.gov/airport/status/airportCode>

Response Formats:
xml, json

HTTP Method(s):
GET

Parameters:
All of the following are required:
airportCode. Required. The three letter airport code for which you wish to retrieve data, e.g., "IAD"

Usage Notes:
• If the system serving the data is unavailable, an HTTP status code of 502 (Bad Gateway) will be returned along with the error message.

Sample XML Request:
<http://services.faa.gov/airport/status/SFO?format=application/xml> (<http://services.faa.gov/airport/status/SFO?format=application/xml>)

Sample XML Response:
[view plaintext](#)

1. <AirportStatus>
2. <!-- Airport Name -->
3. <Name>San Francisco Intl</Name>
4. <!-- International Civil Aviation Organization Airport Code -->
5. <ICAO>KSFO</ICAO>
6. <!-- International Association of Travel Agents Airport Code -->
7. <IATA>SFO</IATA>
8. <!-- Is there a delay? -->
9. <Delay>true</Delay>
10. <!-- Status -->
11. <Status>
12. <!-- Types: Airport Closure, Ground Stop, Ground Delay, Arrival and/or Departure -->
13. <Type>Ground Delay</Type>
14. <!-- For use with All types -->
15. <Reason>LOW CEILING</Reason>
16. <!-- For use with Ground Delay -->
17. <AvgDelay>32 minutes</AvgDelay>
18. <!-- For use with Airport Closures -->
19. <ClosureEnd/>
20. <ClosureBegin/>
21. <!-- For use with Arrival and/or Departure Delays -->
22. <MinDelay/>
23. <Trend/>
24. <MaxDelay/>
25. <!-- For use with Ground Stops -->
26. <EndTime/>
27. </Status>
28. <!-- Weather from NOAA -->
29. <Weather>
30. <Weather>Mostly Cloudy</Weather>
31. <Meta>
32. <Credit>NOAA's National Weather Service</Credit>
33. <Url><http://weather.gov/></Url>
34. <!-- Update Time is local to airport itself -->
35. <Updated>11:56 AM Local</Updated>
36. </Meta>
37. <Wind>South at 9.2mph</Wind>

Figure 71: FAA Web Services, Airport Service home page.

9. Federal Reserve Bank of St. Louis, Economic Research: Download, graph, and track 294,000 US and international time series from 81 sources. (see Figure 72)
<https://research.stlouisfed.org/fred2/>

Browse data by [Tag \(/tags\)](#), [Category \(/categories\)](#), [Release \(/releases\)](#), [Source \(/sources\)](#), [Release Calendar \(/releases/calendar\)](#) or [Get Help \(/help\)](#)

FRED News [\(https://research.stlouisfed.org/rss/\)](https://research.stlouisfed.org/rss/)
 FRED Adds 127 Macroeconomic Series (<http://news.research.stlouisfed.org/?p=2790>)
 FRED Adds 987 World Development Indicators Series (<http://news.research.stlouisfed.org/?p=2786>)
FRED Blog (<https://fredblog.stlouisfed.org/>) [\(https://research.stlouisfed.org/rss/\)](https://research.stlouisfed.org/rss/)
 TED on FRED (<http://fredblog.stlouisfed.org/?p=2515>)
Research News [\(https://research.stlouisfed.org/rss/\)](https://research.stlouisfed.org/rss/)
 Who Benefits from Too Big To Fail? (<http://news.research.stlouisfed.org/?p=2756>)

AT A GLANCE | POPULAR SERIES | LATEST RELEASES | TOOLS | NEED HELP?

CPI (/series/CPIAUCSL) [\(https://research.stlouisfed.org/rss/\)](https://research.stlouisfed.org/rss/)
+1.1 % Chg. from Yr. Ago on May 2016

Real GDP (/series/A191RL1Q225SBEA)
1.1 % Chg. from Preceding Period on Q1 2016

IP (/series/INDPRO)
-0.4 % Chg. on May 2016

10-Yr. Treas. Rate (/series/DGS10)
1.40 % on 2016-07-07

US/Euro FX Rate (/series/DEXUSEU)
1.1145 U.S. \$ to 1 Euro on 2016-07-01

Civ. Unemploy. Rate (/series/UNRATE)
4.9 % on Jun 2016

Payroll Employment (/series/PAYEMS)
+287 Chg., Thous. of Persons on Jun 2016

Initial Jobless Claims, 4-Week Moving Average (/series/IC4WSA)
264750 on 2016-07-02

Federal Reserve Bank of St. Louis, One Federal Reserve Bank Plaza, St. Louis, MO 63102

[Send feedback](#)

Your eyes do not deceive you. FRED's got a new look and features. [Learn more \(https://news.research.stlouisfed.org/2016/06/fred-has-moved-into-fancy-new-digs-take-a-tour/\)](https://news.research.stlouisfed.org/2016/06/fred-has-moved-into-fancy-new-digs-take-a-tour/)

Figure 72: Federal Reserve Bank of St. Louis, Economic Research home page.

10. Gapminder Data: Gapminder is a non-profit venture promoting sustainable global development and achievement of the United Nations Millennium Development Goals by increased use and understanding of statistics and other information about social, economic and environmental development at local, national and global levels. (see Figure 73)
<http://www.gapminder.org/data/>

The screenshot shows the 'Data in Gapminder World' page. At the top, there is a navigation bar with links for 'GAPMINDER', 'GAPMINDER WORLD', 'VIDEOS', 'DOWNLOADS', 'TEACH', 'IGNORANCE', and 'DATA'. Below the navigation bar, the page title is 'Data in Gapminder World'. There are several links: 'List of indicators', 'About countries & territories', 'Documentation', and 'Data blog'. A paragraph explains that the table lists all indicators displayed in Gapminder World and provides instructions on how to access information about the indicator and a link to the data provider. It also notes that indicators labeled 'Various sources' are compiled by Gapminder and can be reused freely with attribution. Below this is the 'List of indicators in Gapminder World' section, which includes a search bar and a table of indicators.

Indicator name	Data provider	Category	Subcategory	Download	View	Visualize
Adults with HIV (% age 15-49)	Based on UNAIDS	Health	HIV			
Age at 1st marriage (women)	Various sources	Population				
Aged 15+ employment rate (%)	International Labour Organization	Work	Employment rate			
Aged 15+ labour force participation rate (%)	International Labour Organization	Work	Labour force participation			
Aged 15+ unemployment rate (%)	International Labour Organization	Work	Unemployment			
Aged 15-24 employment rate (%)	International Labour Organization	Work	Employment rate			
Aged 15-24 unemployment rate (%)	International Labour Organization	Work	Unemployment			
Aged 15-64 labour force participation rate (%)	International Labour Organization	Work	Labour force participation			
Aged 25-54 labour force participation rate (%)	International Labour Organization	Work	Labour force participation			
Aged 25-54 unemployment rate (%)	International Labour Organization	Work	Unemployment			
Aged 55+ unemployment rate (%)	International Labour Organization	Work	Unemployment			
Aged 65+ labour force participation rate (%)	International Labour Organization	Work	Labour force participation			
Agricultural land (% of land area)	World Bank	Environment	Geography			
Agricultural	FAO	Environment	Water			

Figure 73: Gapminder Data home page.

11. General Social Survey: The GSS gathers data on contemporary American society in order to monitor and explain trends and constants in attitudes, behaviors, and attributes. Hundreds of trends have been tracked since 1972. In addition, since the GSS adopted questions from earlier surveys, trends can be followed for up to 70 years. The GSS contains a standard core of demographic, behavioral, and attitudinal questions, plus topics of special interest. Among the topics covered are civil liberties, crime and violence, intergroup tolerance, morality, national spending priorities, psychological well-being, social mobility, and stress and traumatic events. Altogether the GSS is the single best source for sociological and attitudinal trend data covering the United States. It allows researchers to examine the structure and functioning of society in general as well as the role played by relevant subgroups and to compare the United States to other nations. The GSS aims to make high-quality data easily accessible to scholars, students, policy makers, and others, with minimal cost and waiting. (see Figure 74)

<http://www3.norc.org/GSS+Website>

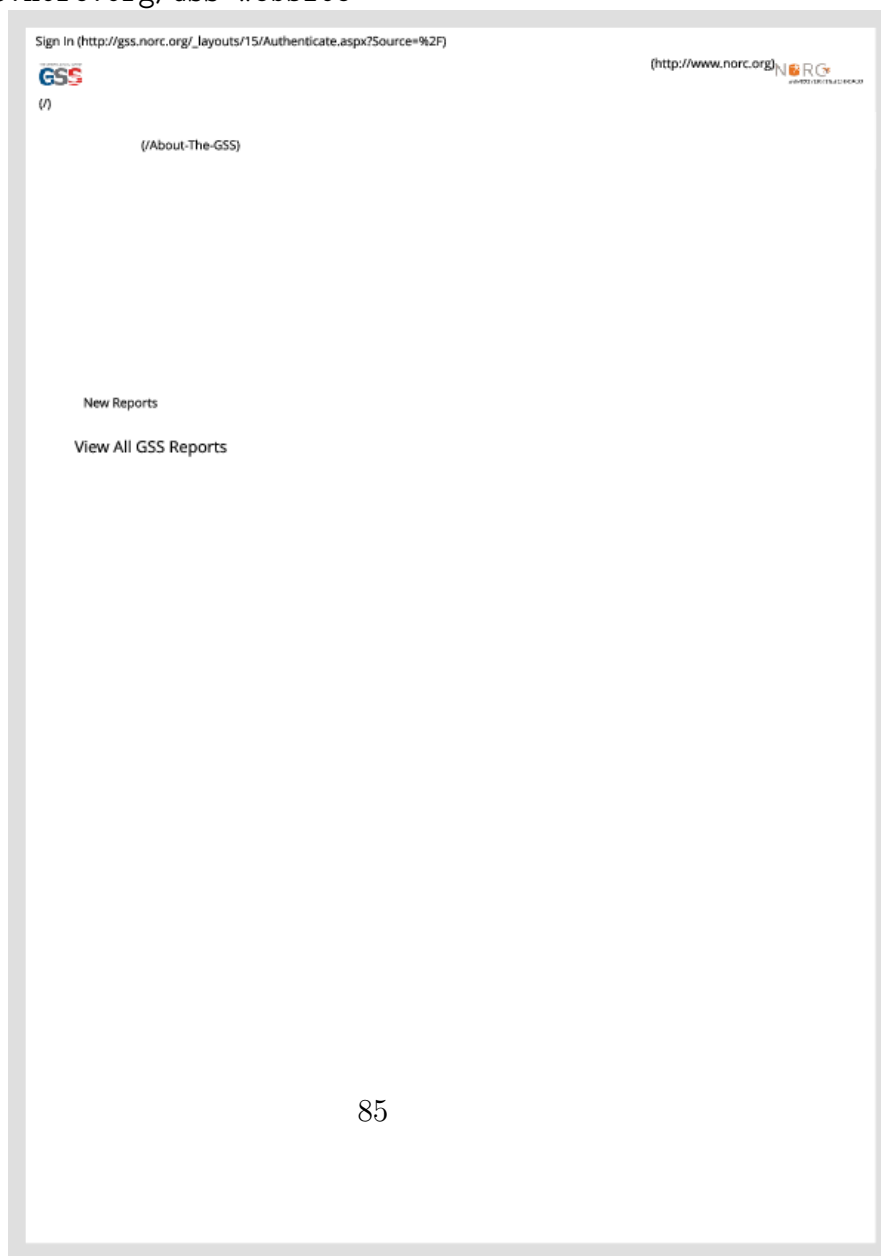


Figure 74: General Social Survey home page.

12. Massachusetts Bay Transportation Authority: Contains schedule, alert, vehicle position, and arrival prediction data. The same data is available in GTFS-realtime. (see Figure 75)
<http://realtime.mbta.com/portal>

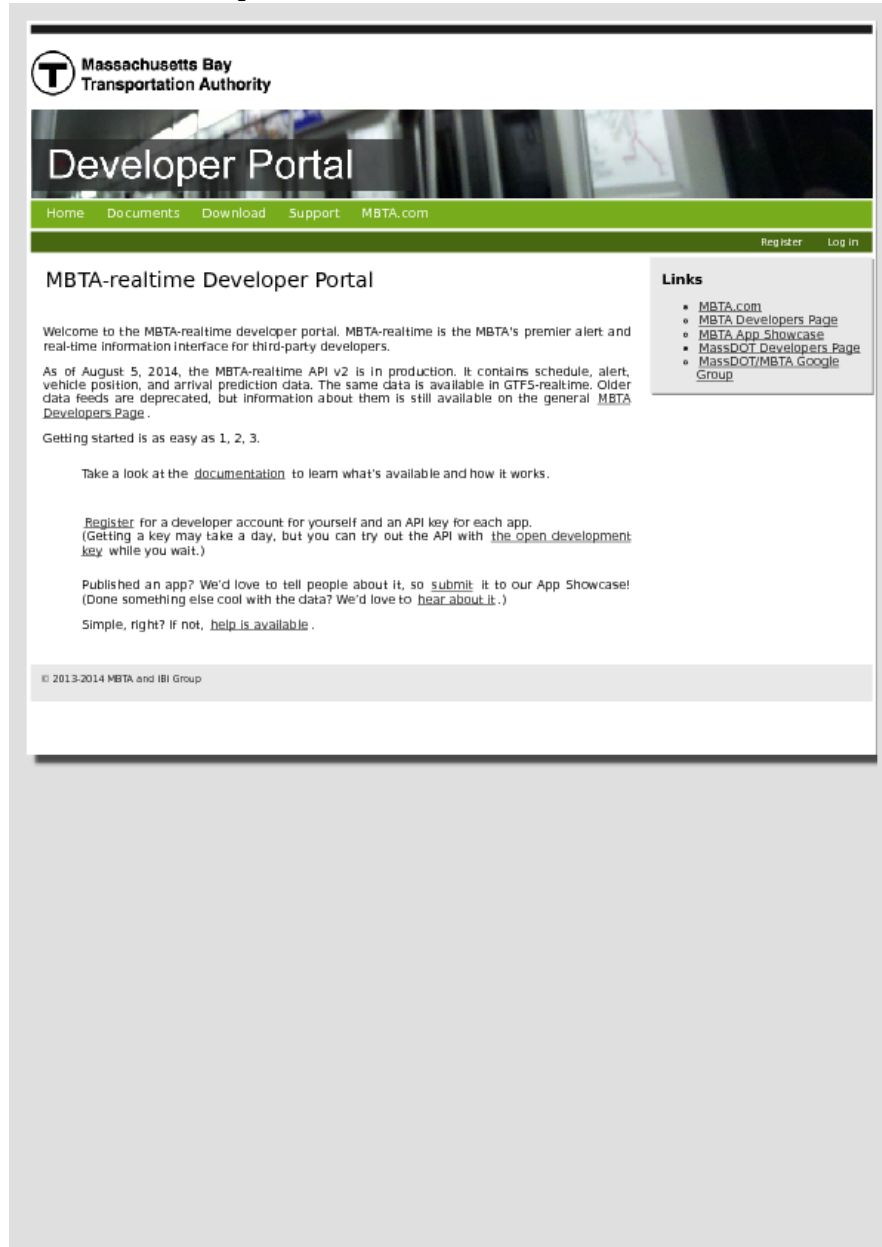


Figure 75: Massachusetts Bay Transportation Authority home page.

13. National Aeronautics and Space Administration, Goddard Institute for Space Studies: Datasets and derived material are available from the NASA GISS websites. (see Figure 76) <http://data.giss.nasa.gov/>



Figure 76: National Aeronautics and Space Administration, Goddard Institute for Space Studies home page.

14. National Flight Data Center (NFDC): The NFDC is part of the FAA's Aeronautical Information Services group (AJV-5). This web portal is intended to service the aviation community providing the ability to browse for aeronautical data, submit data to the FAA for airport or chart updates, or communicate with FAA specialists. (see Figure 77)
<https://nfdc.faa.gov/xwiki/bin/view/NFDC/WebHome>

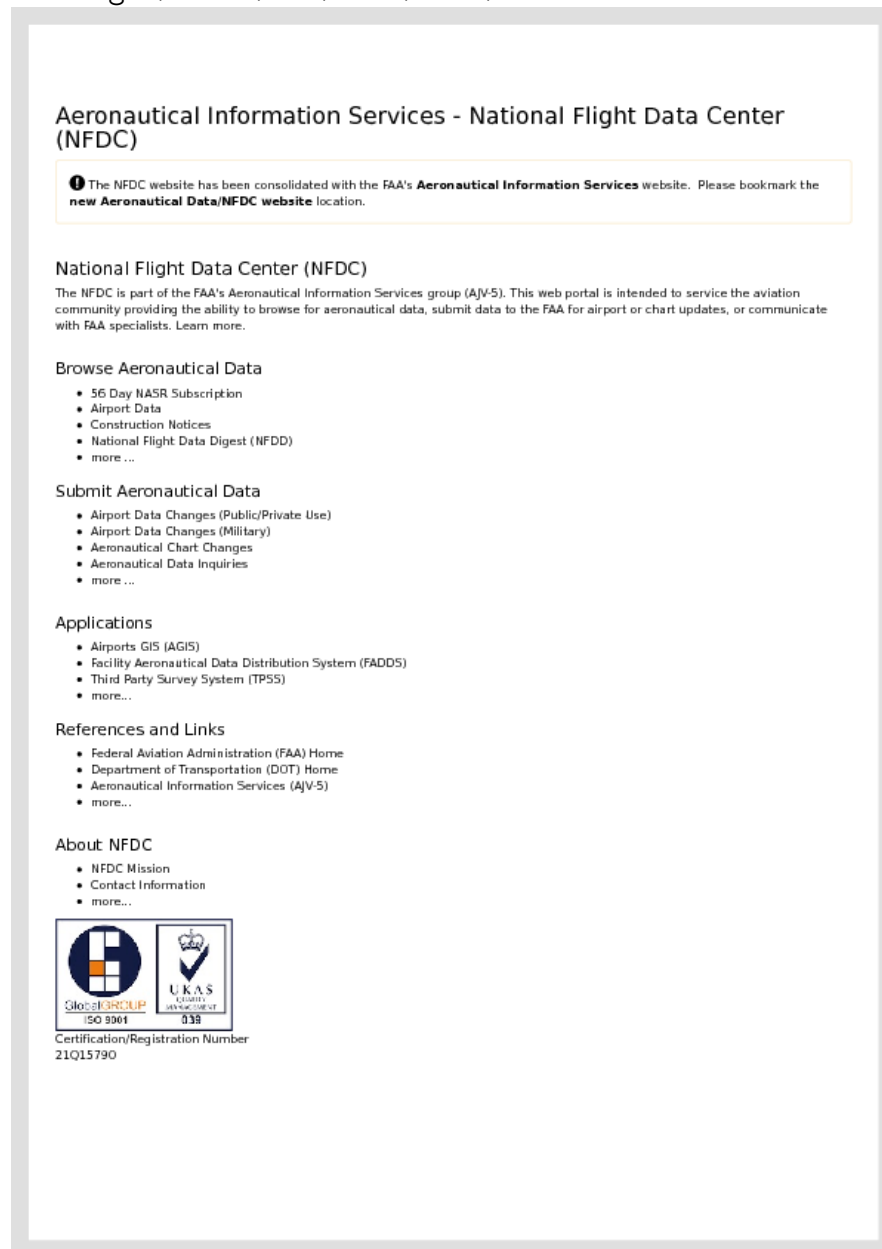


Figure 77: National Flight Data Center (NFDC) home page.

15. National Longitudinal Surveys — A Program of the U.S. Bureau of Labor Statistics: The NLS, sponsored by the U.S. Bureau of Labor Statistics, are nationally representative surveys that follow the same sample of individuals from specific birth cohorts over time. The surveys collect data on labor market activity, schooling, fertility, program participation, health, and much, much more. (see Figure 78)
<https://www.nlsinfo.org/#NLSY79>

National Longitudinal Surveys
 A Program of the U.S. Bureau of Labor Statistics

search site
 Advanced Search

About the National Longitudinal Surveys (NLS)
 The NLS, sponsored by the U.S. Bureau of Labor Statistics, are nationally representative surveys that follow the same sample of individuals from specific birth cohorts over time. The surveys collect data on labor market activity, schooling, fertility, program participation, health, and much, much more. Choose a cohort below to learn more.

- NLS Youth 1997 (NLSY97)**
[/main/en/cohorta/NLSY97](#)
 Men and women born in the years 1980-84
- NLS Mature and Young Women (NLSMW)**
[/main/en/cohorta/0-20years-and-Young-Women](#)
 Mature women born in the years 1922-37 and young women born in the years 1943-53.
- NLS Youth 1979 (NLSY79)**
[/main/en/cohorta/NLSY79](#)
 Men and women born in the years 1957-64
- NLS Older and Young Men (NLSOM)**
[/main/en/cohorta/Older-and-Young-Men](#)
 Older men born in the years 1906-21 and young men born in the years 1941-52.
- NLS 79 Child and Young Adult (NLSY79C)**

Who Are You?
 Please indicate your NLS experience so we can direct you to a page that focuses on your needs:

- New NLS User**
<https://www.nlsinfo.org/content/who-are-you#new-user?>
- Experienced NLS User**
<https://www.nlsinfo.org/content/who-are-you#experienced-user?>

Need Help Using NLS Data?
Explore the Tutorials
<https://www.nlsinfo.org/content/getting-started/intro-to-the-nls/tutorials>
 Learn how to approach research projects, search for and extract data, and then program with NLS data.

Introduction to the Investigator
https://www.nlsinfo.org/investigator/Guide/investigator_guide_TOC.html
 Learn how to use the online NLS data search and extraction site to create your own data sets.

Latest Information from the NLS
Newest Data Release [6/15/2016]: Preliminary 2014 NLSY79 Young Adult Data. A preliminary version of the 2014 NLSY79 Young Adult data is now available to the public through NLS Investigator (<https://www.nlsinfo.org/investigator>) (<https://www.nlsinfo.org/investigator>). The young adults surveyed in this round (age range 14 to 42) are the older children of female respondents in the NLSY79. The release includes all data collected for the YA in the 2014 data collection round, plus a limited number of created variables. These data can be merged in with the NLSY79 Child/YA 1979-2012 data release. Users should be aware that the reference numbers in this preliminary release are not final and may be different in the final merged NLSY79 Child/YA 1979-2014 data release. Question names, however, will not change.

Attending the ASA meeting in Seattle this year? Join us at our NLS workshop there. We're offering an NLS user workshop at ASA on Saturday, August 20 (check ASA program for workshop time and location). Here's your chance to get the latest NLS information and developments, hear an overview on the various NLS datasets, and receive how-to instructions for searching and downloading NLS data. We'll also have our usual NLS exhibit booth at the conference.

More content to come shortly. NLS Resources. Need some quick info about the NLS?

Figure 78: National Longitudinal Surveys — A Program of the U.S. Bureau of Labor Statistics home page.

16. NYC Open Data: Open data government data about New York state. (see Figure 79)
<https://data.cityofnewyork.us/>

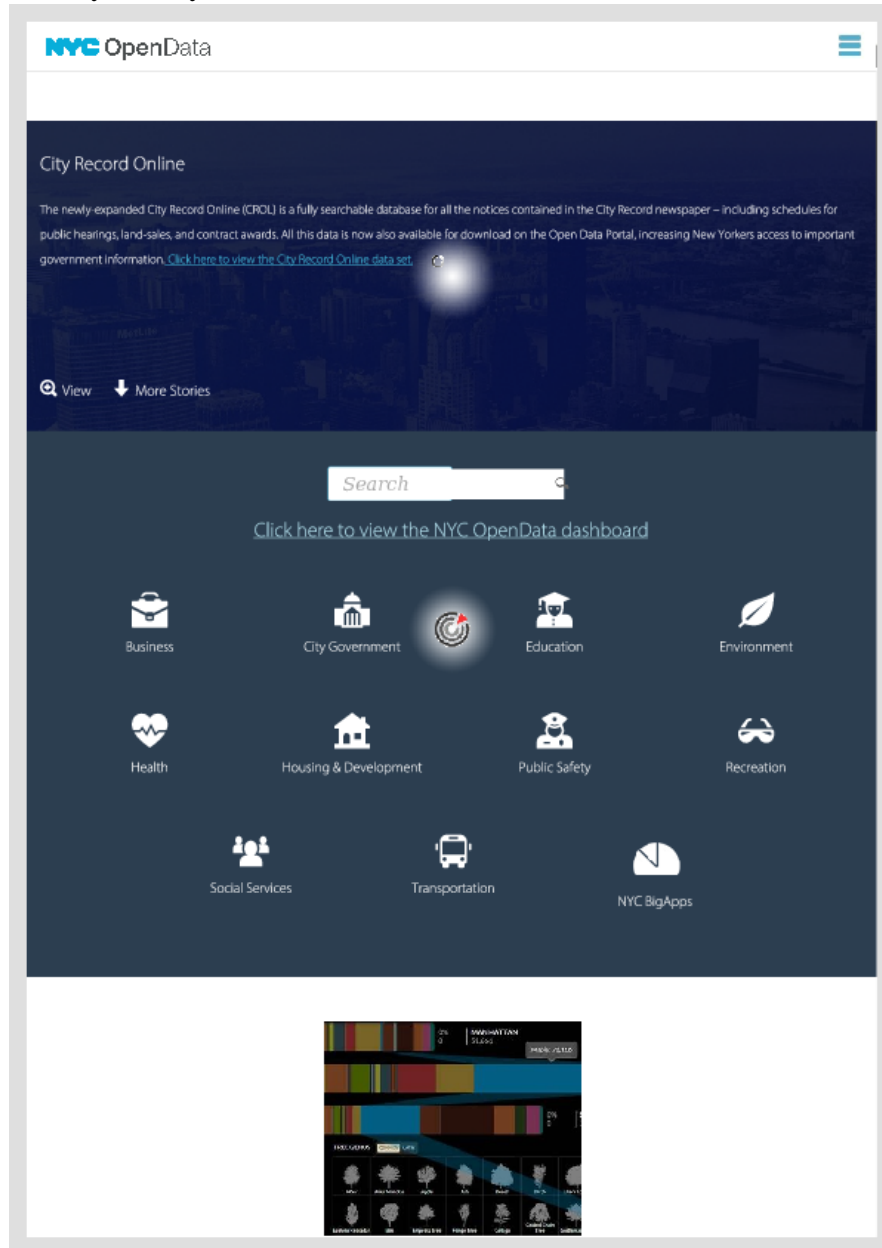


Figure 79: NYC Open Data home page.

17. Parking in San Francisco, CA: The PDF contains the URL to the API that provides real-time parking information for San Francisco, CA. (see Figure 80)
<http://sfpark.org/resources/sfpark-service-api-reference-updated/>

The screenshot shows the SFpark website home page. At the top right, there are links for "Skip to Main Content", "Font Size", "Web Accessibility", and "Video Transcript". The SFpark logo is prominently displayed on the left. Below the logo is a navigation menu with categories like "About", "How it Works", "Resources", "News", and "Contact". A search bar is located below the menu. The main content area features a "Resources" section with a "Next Resource" and "Prev Resource" link, and a featured article titled "SFpark service API reference - UPDATED". To the right of the main content are two sidebars: "Resource Types" and "Resource Categories", both listing various resource counts. At the bottom, there is a "Share This Page" section with social media icons and a copyright notice for 2016.

Figure 80: Parking in San Francisco, CA home page.

18. SNAP Retailer Locator: The location of retailers that take/honor Supplemental Nutrition Assistance Program (SNAP) cards. (see Figure 81)
<http://www.fns.usda.gov/snap/retailerlocator>

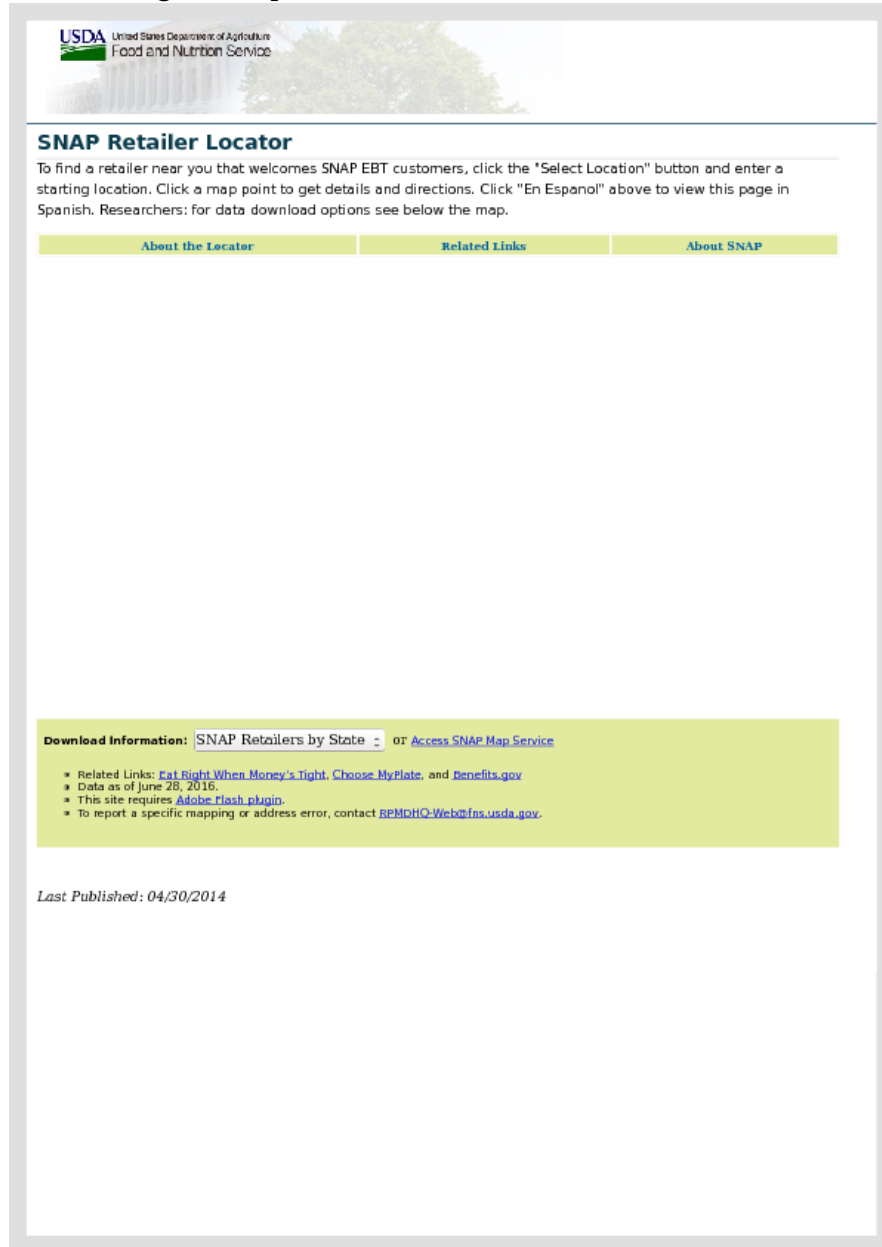


Figure 81: SNAP Retailer Locator home page.

19. Social Security Administration: Welcome to the data page of the Social Security Administration. On this page there are links to resources related to agency data resources and to the federal government policies and direction for opening data to the public. We have been proactive in releasing data to the public because government data has a strategic value. (see Figure 82)
<https://www.ssa.gov/data/>

Social Security

Social Security Data Page

Welcome to the data page of the Social Security Administration. On this page there are links to resources related to agency data resources and to the federal government policies and direction for opening data to the public. We have been proactive in releasing data to the public because government data has a strategic value. Citizens, businesses, and other governmental and non-profit organizations need data and its related information for:

- Greater transparency and accountability
- Understanding mission and operations
- Economic growth
- Innovation and research
- Education and training
- Making decisions about their lives, businesses, and organizations

The May 9, 2013 Executive Order, Making Open and Machine Readable the New Default for Government Information, establishes the overall direction all federal agencies, including Social Security, must take in moving to an open data environment. The purpose of the order is to make "information resources easy to find, accessible, and usable." The premise is that human readable text is not sufficient for many of the innovative uses of data and that opening up the government's data resources for innovation is best done using machine readable formats and open platforms. Following the executive order, the Office of Management and Budget issued M-13-13, Open Data Policy-Managing Information as an Asset, which provides more detailed agency requirements for implementing open data. The directive explicitly requires agencies to use machine-readable and open formats. All data released to the public may be used without restriction.

In addition to detailing open and accessible requirements, M-13-13 directs agencies to create and maintain an enterprise data inventory and to identify in a public data listing those data resources that have or can be released to the public. Agencies must also create a process to engage with customers to help facilitate and prioritize data releases. All of the open data activities must be carefully performed to ensure that privacy and confidentiality are fully protected and that data are properly secured.

Social Security's public data listing in machine readable format is located at the following address (www.ssa.gov/data.json). Below is the human readable version with filtering and links to additional content where applicable. We will continue to expand the contents of our public data listing regularly as new information becomes available.

We will continue to solicit ideas from our stakeholders and we will hold public feedback engagements to better understand your needs and to explain the data we make available. Some of engagements will be in person and others will use automated tools, like IdeaScale, to help organize and track ideas from the public. These conversations will help to determine our priorities in terms of specific data that we should be releasing and other efforts we should be employing to support open and transparent access to our data assets. You can share your ideas about our open data program using SSA's IdeaScale community, or submit your ideas using our feedback form. We invite you to review our Customer Feedback page for information about some of our data customers' key ideas and how we plan to address them.

Access Level	Name	Description	Downloads

Share Your Ideas

SSA's IdeaScale community for Open Data Project

SSA's Open Data Themes

Resources and links

Executive Order, Making Open and Machine Readable the New Default for Government Information
M-13-13, Open Data Policy-Managing Information as an Asset

Figure 82: Social Security Administration home page.

20. The National Archives: The United Kingdom National Archives. (see Figure 83)
<http://www.nationalarchives.gov.uk/>

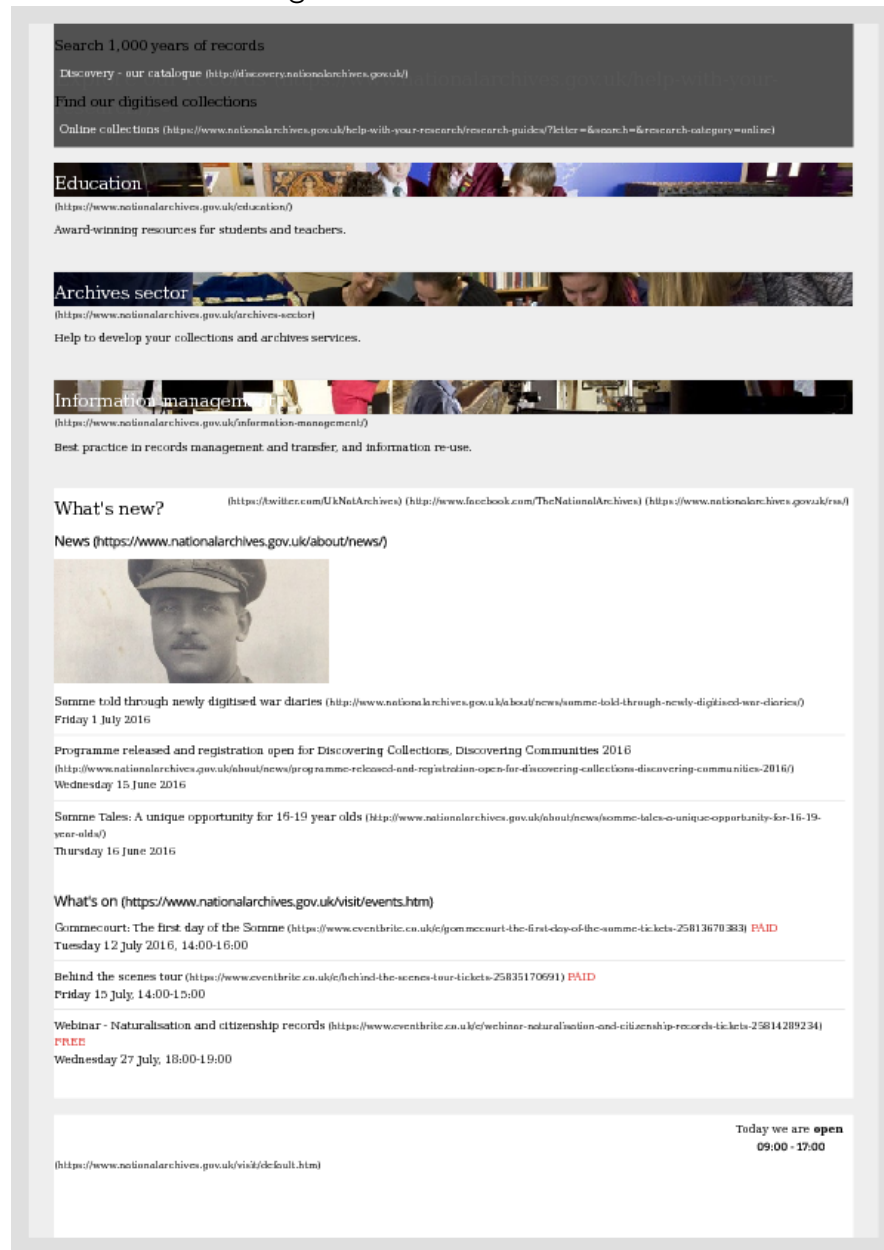


Figure 83: The National Archives home page.

21. The National Map: As one of the cornerstones of the U.S. Geological Survey's (USGS) National Geospatial Program, The National Map is a collaborative effort among the USGS and other Federal, State, and local partners to improve and deliver topographic information for the Nation. (see Figure 84)
<http://nationalmap.gov/>



Figure 84: The National Map home page.

22. The National UFO Reporting Center: Dedicated to the Collection and Dissemination of Objective UFO Data (see Figure 85)
<http://www.nuforc.org/webreports.html>

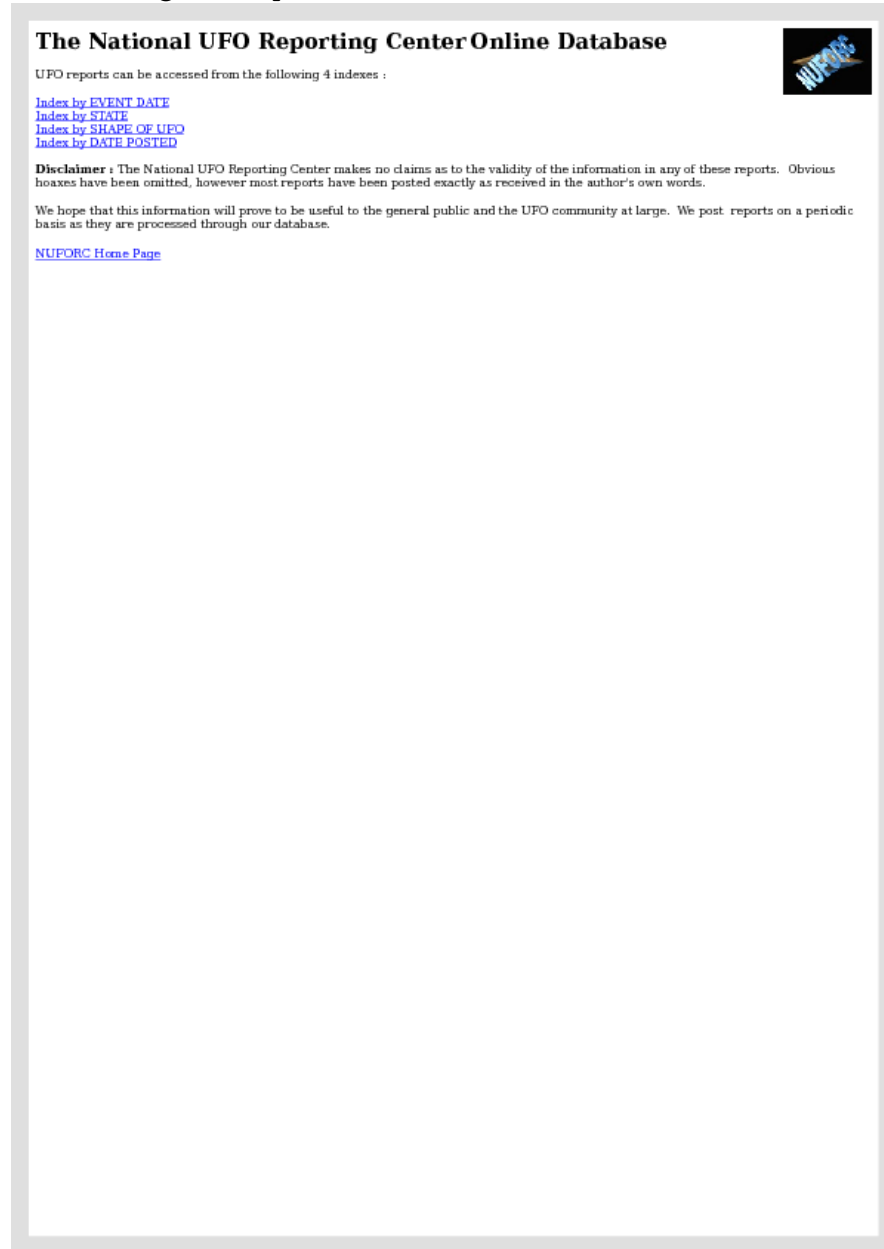


Figure 85: The National UFO Reporting Center home page.

23. The World Bank: These datasets were compiled for World Bank research and are provided free of cost to foster the creation of new knowledge. (see Figure 86)
<http://www.worldbank.org/>



Figure 86: The World Bank home page.

24. UK Government Open Data Portal: U.S. Government open data portal. (see Figure 87)
<https://data.gov.uk/>



Figure 87: UK Government Open Data Portal home page.

25. United States Department of Agriculture, Economic Research Service: The ERS mission is to inform and enhance public and private decision making on economic and policy issues related to agriculture, food, the environment, and rural development. With over 300 employees, The Economic Research Service is a primary source of economic information and research in the U.S. Department of Agriculture. (see Figure 88)
<http://www.ers.usda.gov/data-products/.aspx>

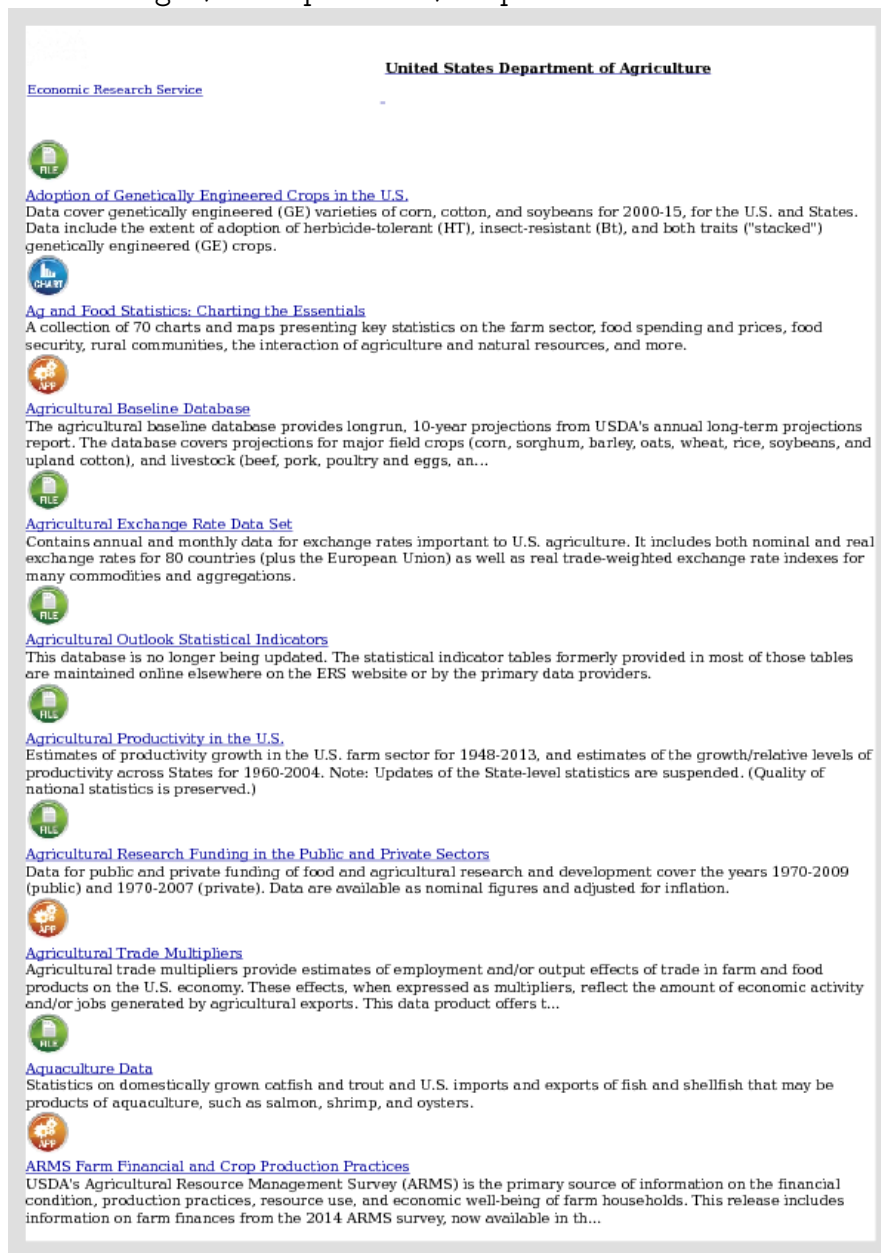


Figure 88: United States Department of Agriculture, Economic Research Service home page.

26. USDA Food Composition Database: Access to Release 28 of the USDA National Nutrient Database for Standard Reference. You can either view the data here or download the data files and documentation in several different formats for use on your computer. An online search is also provided so you can look up the nutrient content of 8,789 different foods directly from this home page. (see Figure 89)
<https://ndb.nal.usda.gov/>

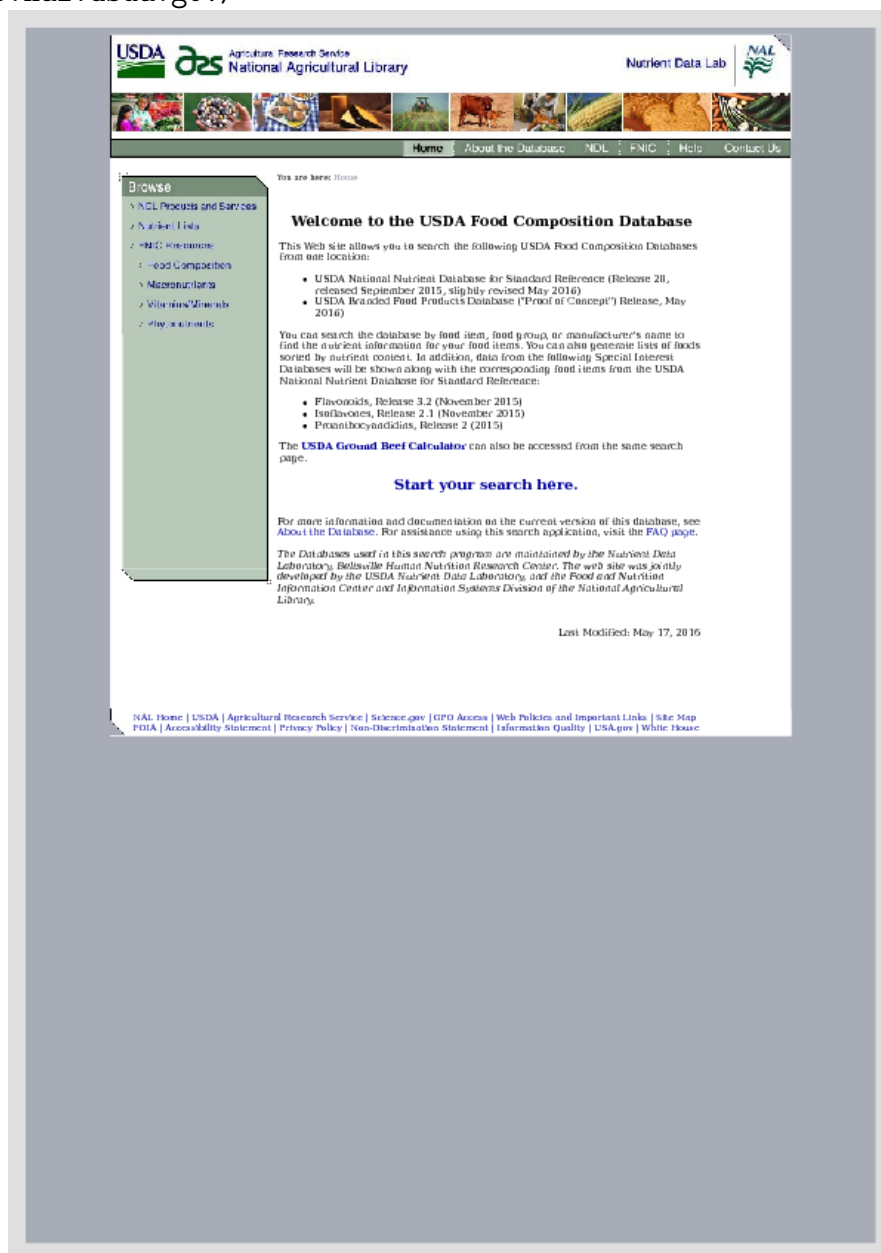


Figure 89: USDA Food Composition Database home page.

27. US Department of Transportation: Catalog of available datasets. (see Figure 90)
<https://catalog.data.gov/organization/dot-gov>

The screenshot shows the Data.gov profile for the US Department of Transportation. At the top, there is a search bar and navigation links for DATA, TOPICS, IMPACT, APPLICATIONS, DEVELOPERS, and CONTACT. Below this is the 'DATA CATALOG' section with a search bar and a dropdown menu for 'Order by'. The main content area displays '812 datasets found' and lists several datasets. The first dataset is 'Military Bases (National)' with 56 recent views. The second is 'Airline On-Time Performance and Causes of Flight Delays' with 42 recent views. The third is 'SAFER - Company Snapshot' with 31 recent views. The fourth is 'National Bridge Inventory - National Geospatial Data Asset (NGDA) Bridges' with 23 recent views. The fifth is 'National Highway Planning Network (NHPN)' with 23 recent views. On the left side, there is a profile for the 'Department of Transportation' with 0 followers and 1 member. Below the profile are sections for 'Topics' and 'Topic Categories' with filters and search options.

Figure 90: US Department of Transportation home page.

28. US Government Open Data Portal: U.S. Government open data portal to over 190,000 datasets. (see Figure 91)
<http://www.data.gov/>

The screenshot shows the homepage of the US Government Open Data Portal. At the top, it says "The home of the U.S. Government's open data" and "Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and [more \(/about\)](#)." Below this is a "GET STARTED" section with a search bar containing "Monthly House Price Indexes" and a search icon. A "BROWSE TOPICS" section follows, featuring a grid of icons and labels for various categories: Agriculture (/food), Business (/business), Climate (/climate), Consumer (/consumer), Ecosystems (/ecosystems), Education (/education), Energy (/energy), Finance (/finance), Health (/health), Local Government (/local), Manufacturing (/manufacturing), Ocean (/ocean), Public Safety (/safety), and Science & Research (/research). At the bottom, there is a "HIGHLIGHTS" section titled "7 Years of Open Data" with a paragraph of text and several links to related content.

Figure 91: US Government Open Data Portal home page.

29. USGovXML.com: US Government Web Services and XML Data Sources.: USGovXML.com is an index to publicly available web services and XML data sources that are provided by the US government. USGovXML.com indexes data sources from all 3 branches of government as well as its boards, commissions, corporations and independent agencies. (see Figure 92) <http://usgovxml.com/>

US Government Web Services and XML Data Sources

Search USGovXML

Home Index About Contact Examples Help Mobile Apps

US Government Data Sources

- HOME
- Boards
 - Federal Reserve Board
 - Recovery Board
- Commissions
 - Consumer Product Safety
 - Federal Communications
 - Federal Election
 - Federal Trade
 - International Trade
 - National Capital Planning
 - Securities and Exchange
- Councils
 - Financial Institutions Exam
- Departments
 - Agriculture
 - Commerce
 - Defense
 - Energy
 - Health & Human Services
 - Homeland Security
 - Housing and Urban Dev.
 - Interior
 - Justice
 - Labor
 - State
 - Transportation
 - Treasury
 - Veterans Affairs
 - Executive Office
- OSTP
- Government Corporations
 - Integ. Technomic Info. Sys.
 - PRBC
 - TVA
- Independent Agencies
 - Environmental Prot. Agency
 - Ex-Im Bank
 - General Services Admin
 - Labor Relations Authority
 - NASA
 - National Archives
 - Natl Science Foundation
 - NTSB
 - Office of Mgmt. and Budget
 - Office of Personnel Mgmt.
 - Peace Corps
 - Selective Service System
 - Small Business Admin
 - Social Security Admin
 - US Postal Service
- US Congress
 - House of Representatives
 - Library of Congress
 - Senate
 - US Courts

US Government Web Services and XML Data Sources

Welcome to USGovXML.com. USGovXML.com is an index to publicly available web services and XML data sources that are provided by the US government. USGovXML.com indexes data sources from all 3 branches of government as well as its Boards, commissions, corporations and independent agencies.

Mobile Apps

As a demonstration of the utility of these publicly available data sources, some of them have been used to create mobile web apps which can be installed on your mobile device or accessed directly from the web browser of any tablet or smart phone! The apps are **free** and are available at the Google Chrome Web Store, the Firefox Marketplace and the Amazon.com App Store.

Description	Chrome Windows, OS X, Linux, Chromebook and mobile devices (via Cordova)	Firefox Firefox OS	Amazon Android, Fire TV, Fire Phone, and Fire Tablets
Mobile Drug Guide *** Amazon Best Seller *** Provides up to date consumer and clinical information about drugs in the nation's Prescription formulary. Information is obtained from prominent US Federal data sources and includes drug monographs, clinical trials, drug-drug interactions, active ingredients, pharmacological classes, current and recent drug shortages, DEA schedules, recalls, images, manufacturer information, etc. Direct Link: http://mdg.usgovxml.com			
USGovXML Apps An eclectic mix of mobile web apps for outdoor enthusiasts, environmentalists, volunteers and more ... Get current weather and forecasts for any location within the US; Get information about federal recreational facilities, threatened and endangered species, federally supported health centers, and more ... Direct Link: http://m.usgovxml.com			
Alternative Fuels Station Locator Search for alternative fueling stations across the United States. Alternative fuels include biodiesel, compressed natural gas (CNG), ethanol (E85), electric charging, hydrogen, liquefied natural gas (LNG) and propane (LPG). Direct Link: http://afsl.usgovxml.com			
Vehicle Safety Recalls Search the National Highway Transit Safety Administration (NHTSA) database for vehicle safety recalls. Direct Link: http://vsr.usgovxml.com			
Five Star Safety Ratings Provides detailed information about the results of controlled crash and roll over tests conducted on new cars at NHTSA research facilities. Direct Link: http://fsr.usgovxml.com			
HIV/AIDS Mobile Drug Guide Provides access to the latest, federally approved HIV/AIDS medical practice guidelines, information on HIV/AIDS treatment, clinical			

Envision beautiful science

Find out more

invitrogen
by ThermoFisher Scientific

Figure 92: USGovXML.com: US Government Web Services and XML Data Sources. home page.

30. Washington Metropolitan Area Transit Authority API: API with real time information about bus, train, and stations in the Washington, DC Metro system. (see Figure 93)
<https://developer.wmata.com/>

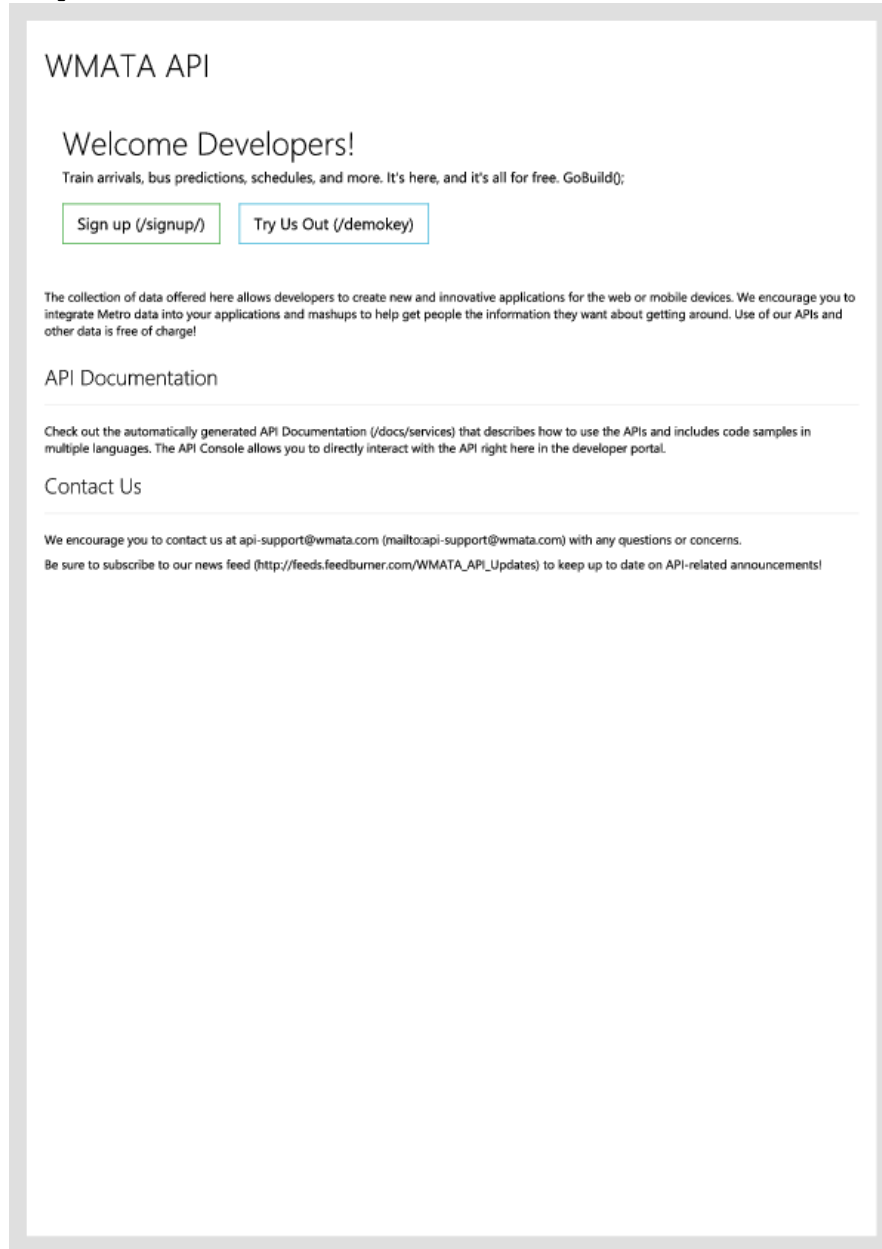


Figure 93: Washington Metropolitan Area Transit Authority API home page.

3.9 Social

1. American Customer Satisfaction Index: The American Customer Satisfaction Index provides unique customer experience benchmarking capabilities that come from the Index's one-of-a-kind, cross-industry structure. (see Figure 95)
<http://theacsi.org/>



Figure 94: American Customer Satisfaction Index home page.

2. American Customer Satisfaction Index: The American Customer Satisfaction Index provides unique customer experience benchmarking capabilities that come from the Index's one-of-a-kind, cross-industry structure. (see Figure 95)

<http://theacsi.org/>

The screenshot shows the ACSI home page with the following elements:

- Header:** ACSI logo with the tagline "American Customer Satisfaction Index". Navigation links include "Contact Us | Client Login", a search bar, and "Newsletter Sign Up".
- Primary Navigation:** Home, Industries, Customer Satisfaction Benchmarks, National Economic Indicator, Products and Services, About ACSI, News and Resources, Global Partnerships.
- Click to show Menu**
- Industry Performance:**
 - FAST FOOD 79 ▲ +2.6% latest results »
 - FULL-SERVICE RESTAURANTS 81 ▼ -1.2% latest results »
 - CELL PHONES 79 ▲ +1.3% latest results »
- Blog Highlights:**
 - Fast Food Ups Its Game:** Patron satisfaction for fast food industry surges as quality improves. **ACSI RESTAURANT REPORT 2016** »
 - New Customer Satisfaction Benchmarks:** **National ACSI Q1 2016** » **Full-Service Restaurants** » **Fast Food** »
 - Blog Highlights: ACSI Matters Smartphones 2016: It's a Galaxy and iPhone Universe:** When it comes to pleasing consumers, the smartphone market is essentially a...
 - Priceline Is Most Appealing Online Booking Site:** Consumers respond well to naming their own price as online travel agency...
 - Would Guests Prefer Starwood Under Marriott Umbrella?:** The merger of Marriott International and Starwood Hotels & Resorts would create...
- ACSI Client Portal:** ACSI's **Client Portal** is the gateway to detailed, confidential results specific to your organization and to peers and competitors across your industry. [Read More »](#)
- ACSI Solutions:**
 - ACSI Solutions:** The **American Customer Satisfaction Index**, the nation's only cross-industry measure of customer satisfaction, gives businesses science-based insights across the complete arc of the customer experience.
 - ACSI BenchmarkSM:** is a total CEX measurement and tracking solution, enabling companies to benchmark all aspects of the customer experience with industry peers and best-in-class companies in other industries.
 - ACSI MonitorSM:** helps clients put the science of ACSI to work within their own customer or employee-based surveys.
 - ACSI DiagnosticSM:** gives companies a powerful tool set for obtaining detailed, actionable insights to improve their customer experience, optimize satisfaction, and maximize results.
- ACSI Around the Globe:** [Learn more about the ACSI Difference »](#)
- ACSI In the News:**
 - CIO: DESPITE IMPROVEMENT, CONSUMERS STILL SAY WIRELESS CARRIERS SUCK** »
 - Yahoo! Finance: A NEW CUSTOMER SERVICE SURVEY SAYS COMCAST NO LONGER THE WORST** »
 - Forbes: AIRLINES' CUSTOMER SERVICE PERFORMANCE HAS IMPROVED, BUT THEY'RE STILL NEAR THE BOTTOM OF THE BARREL** »
 - The Washington Post: WE ALL LOVE TO COMPLAIN ABOUT AIRLINES, BUT CUSTOMER SATISFACTION IS AT AN ALL TIME-HIGH** »
 - Chicago Tribune: RETAIL SURVEY: CUSTOMER SATISFACTION HAS FALLEN SINCE RECESSION 'HALO PERIOD'** »
 - Federal News Radio: SATISFACTION WITH FEDERAL GOVERNMENT SERVICES HITS NEW LOW** »
- [More ACSI media coverage »](#)

Figure 95: American Customer Satisfaction Index home page.

3. LinkedIn API: The foundation of all digital integrations with LinkedIn. (see Figure 96)
<https://developer.linkedin.com/docs/rest-api>

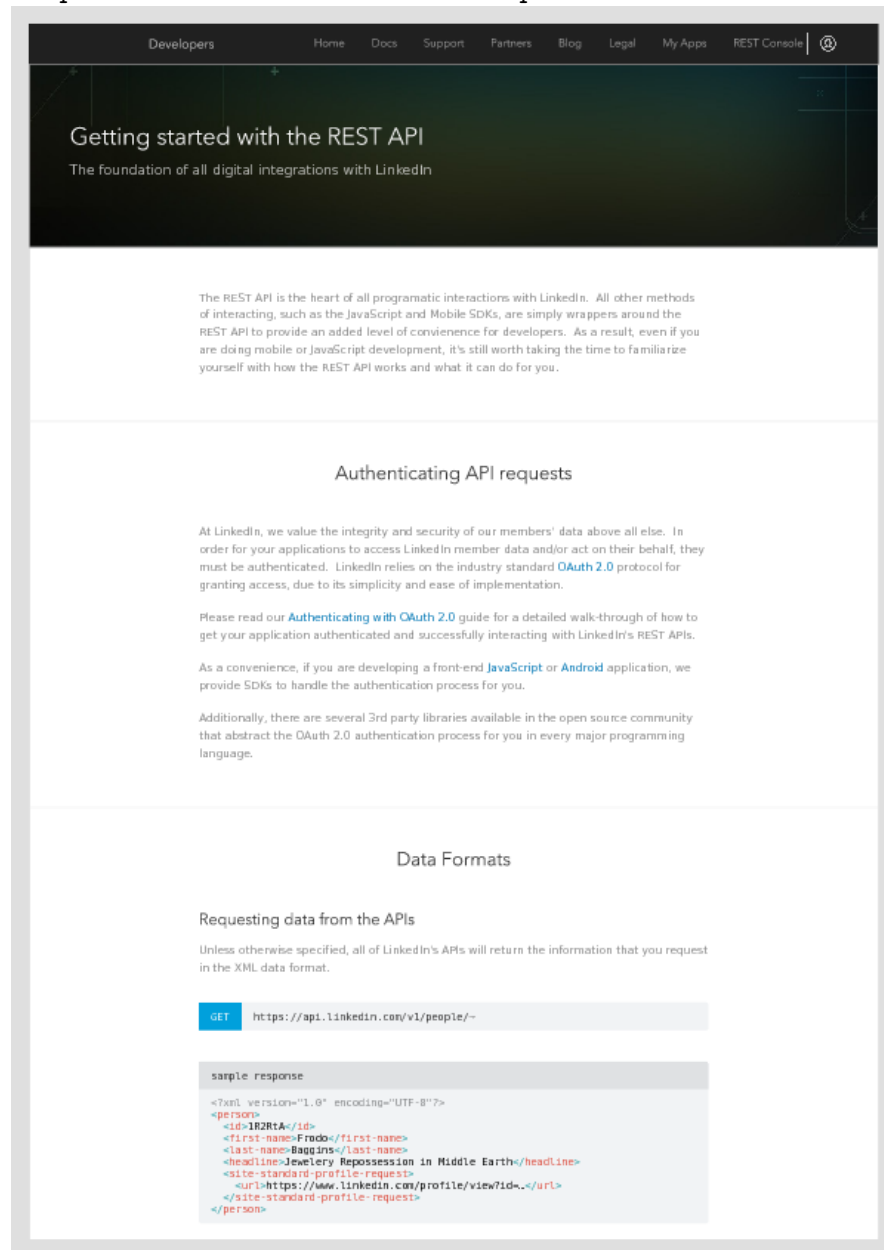


Figure 96: LinkedIn API home page.

4. Meetup API: The Meetup API provides simple RESTful HTTP and streaming interfaces for extending your community using the Meetup platform from your own apps. (see Figure 97) http://www.meetup.com/meetup_api/

The screenshot shows the Meetup API home page. At the top, there's a navigation bar with the Meetup logo and links for 'Log in' and 'Sign up'. Below that, a secondary navigation bar includes 'Meetup API', 'Documentation', 'Console', 'API Key', 'OAuth Consumers', 'Widget Foundry', 'Applications', and 'Client Libraries'. The main content area is titled 'Extend your community' and features a large blue graphic with the Meetup logo and 'API' text. To the right, a sidebar lists 'API methods' with a table of endpoints and methods. The main content is divided into several sections: 'Community' with links to 'Test out your idea', 'Report issues', and 'Join the discussion'; 'Composing a request' with a list of parameters and a sample URL; and 'Cross Origin Resource Sharing' with a brief explanation. The sidebar methods include 'v3 abuse', 'v3 batch', 'v3 boards', 'v2 categories', 'v2 cities', 'v3 comments', 'v2 dashboard', and 'v3 events'.

Figure 97: Meetup API home page.

5. YELP API: Search over 50 million local businesses from 32 countries. Enhance your app with Yelp ratings, reviews, photos and much more. Simple and fast API with powerful category and geo search filters. (see Figure 98)

<https://www.yelp.com/developers>

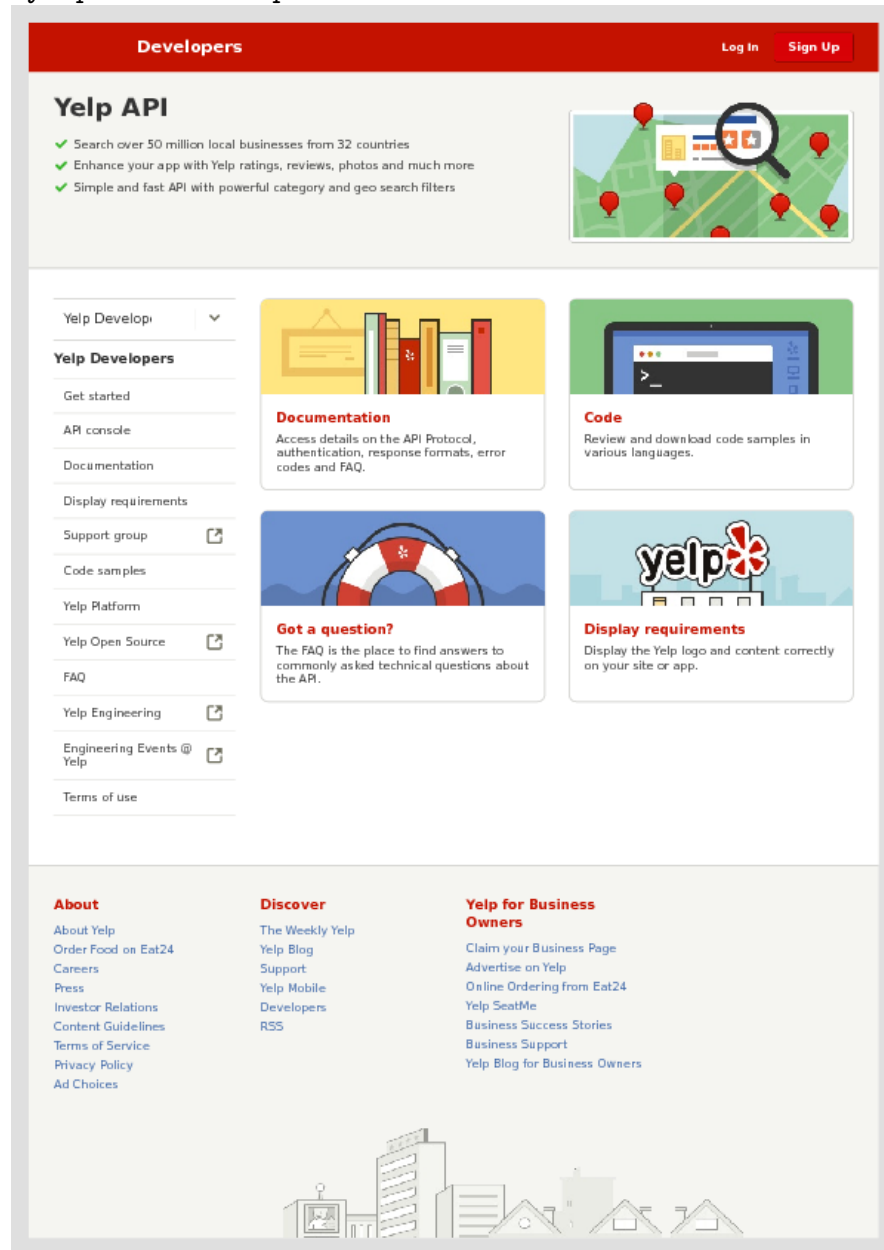


Figure 98: YELP API home page.

3.10 Weather

1. AccuWeather: The AccuWeather API provides subscribers access to location based weather data via a simple RESTful web interface (see Figure 99)
<http://apidev.accuweather.com/developers/>

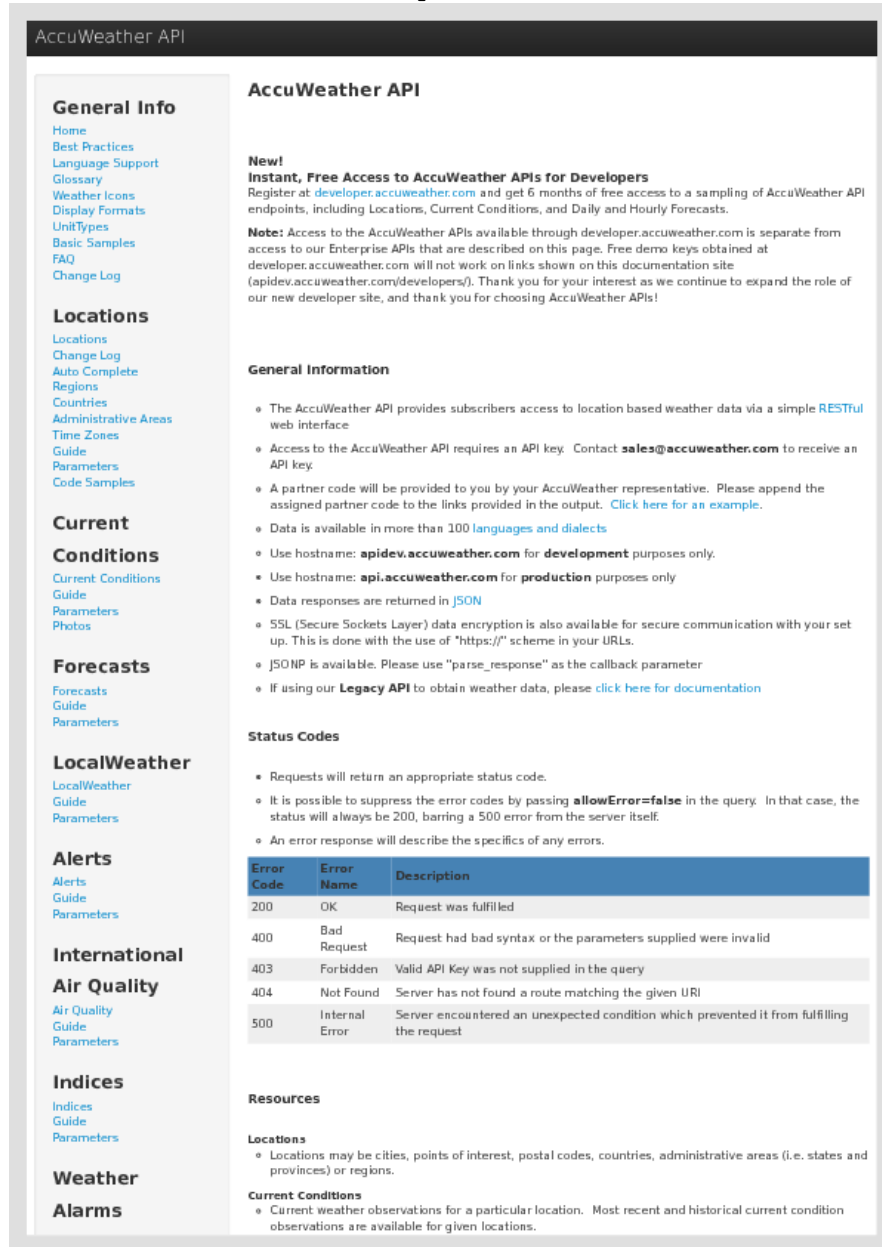


Figure 99: AccuWeather home page.

2. Aeris Weather: An advanced weather API to power all of your custom applications, offering a breath of fresh air from the basic to the most complex solutions. (see Figure 100)
<http://www.aerisweather.com/develop/>

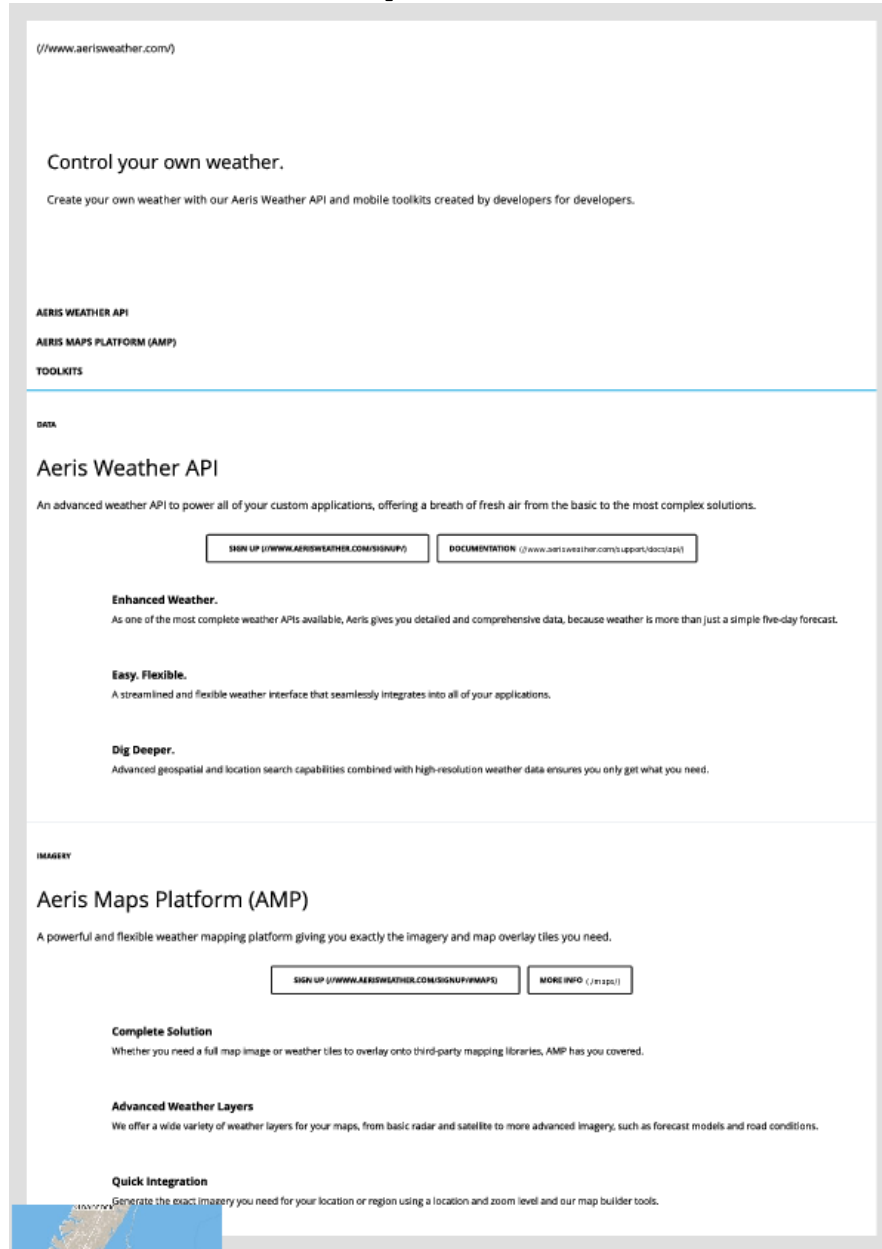


Figure 100: Aeris Weather home page.

3. Dark Sky Forecast: The same API that powers Forecast.io and Dark Sky for iOS can provide accurate shortterm and longterm weather predictions to your business, application, or crazy idea. (see Figure 101)

<https://developer.forecast.io/>



Figure 101: Dark Sky Forecast home page.

4. National Data Buoy Center: The National Data Buoy Center (NDBC) is a part of the National Oceanic and Atmospheric Administration's (NOAA) National Weather Service (NWS). NDBC designs, develops, operates, and maintains a network of data collecting buoys and coastal stations. (see Figure 102)
<http://www.ndbc.noaa.gov/>



Figure 102: National Data Buoy Center home page.

5. National Weather Service: The National Weather Service is a component of the National Oceanic and Atmospheric Administration (NOAA). NOAA is an Operating Unit of the U.S. Department of Commerce. Our Mission Provide weather, water, and climate data, forecasts and warnings for the protection of life and property and enhancement of the national economy. (see Figure 103)
<http://www.weather.gov/>

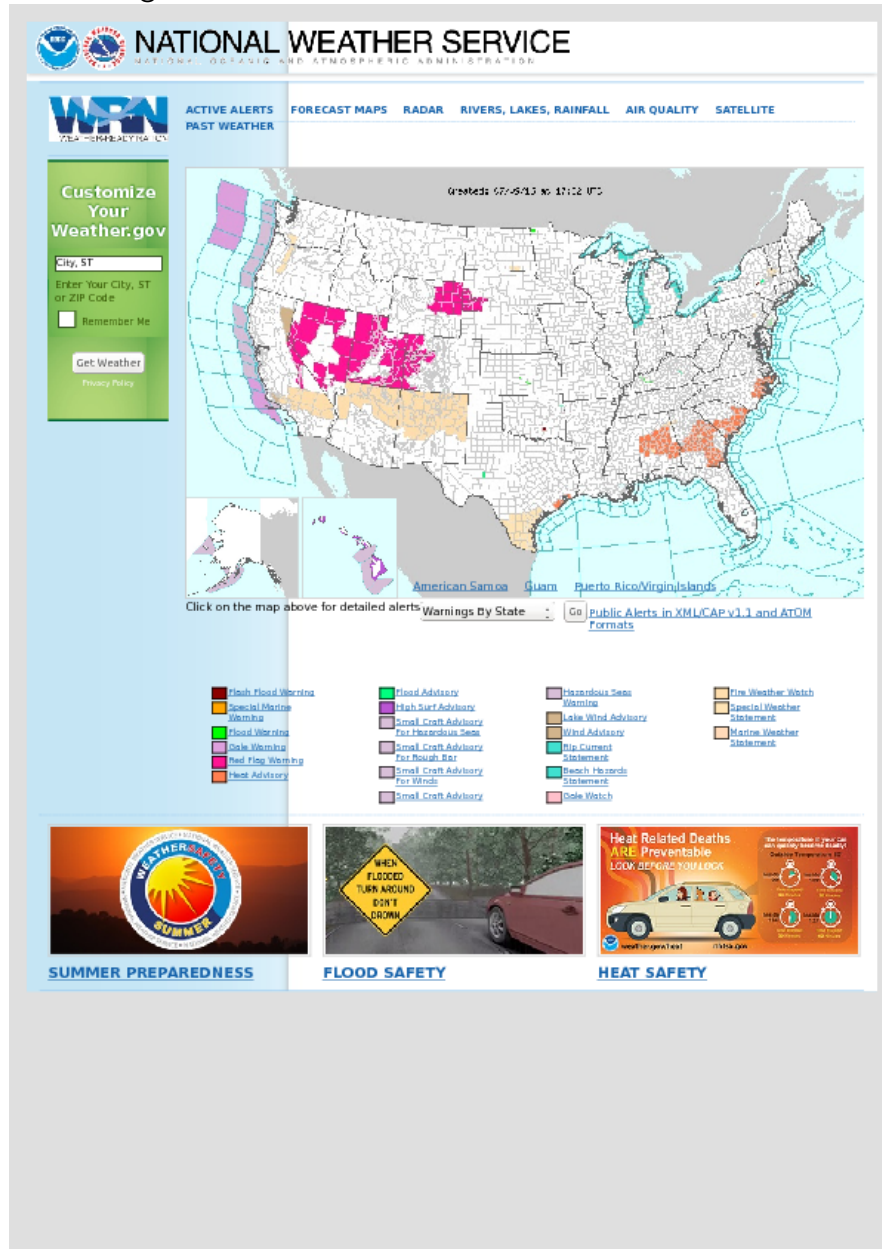


Figure 103: National Weather Service home page.

6. OpenWeatherMap: Our weather API is simple, clear and free. We also offer higher levels of support, please see our paid plan options. To access the API you need to sign up for an API key if you are on a free or paid plan. (see Figure 104)

<http://openweathermap.org/api>

The image shows a screenshot of the OpenWeatherMap API documentation page. The page is titled "Weather API" and contains several sections, each with a heading, a sub-heading, and a list of bullet points. Each section also includes a text input field for an API key. The sections are: "Current weather data", "5 day / 3 hour forecast", "16 day / daily forecast", "Historical data", "UV Index", "Weather map layers", and "Weather stations".

Weather API

Our weather API is simple, clear and free. We also offer higher levels of support, please see our paid plan options. (/price) To access the API you need to sign up for an API key (<http://openweathermap.org/appid>) if you are on a free or paid plan.

Current weather data

API doc (/current)

- Access current weather data for any location including over 200,000 cities
- Current weather is frequently updated based on global models and data from more than 40,000 weather stations
- Data is available in JSON, XML, or HTML format
- Available for Free and all other paid accounts

5 day / 3 hour forecast

API doc (/forecast5)

- 5 day forecast is available at any location or city
- 5 day forecast includes weather data every 3 hours
- Forecast is available in JSON, XML, or HTML format
- Available for Free and all other paid accounts

16 day / daily forecast

API doc (/forecast16)

- 16 day forecast is available at any location or city
- 16 day forecasts includes daily weather
- Forecast is available in JSON, XML, or HTML format
- Available for Developer, Professional and Enterprise accounts

Historical data

API doc (/history)

- Through our API we provide city historical weather data for 20,000+ cities
- Historical data is available for 1 month previous in Starter account, for 1 year previous in Medium accounts, and is 5 and more year previous in Advanced accounts

UV Index

API doc (/api/uv)

- Current UV index (Clear Sky) and historical data are available for any geo location (lat/lon)
- Interpreting of the UV Index and recommended protection are provided
- Data is available in JSON
- Available for Professional and Enterprise accounts

Weather map layers

API doc (/hugemaps)

- Weather maps include precipitation, clouds, pressure, temperature, wind, and more
- Connect our weather maps to your mobile applications and websites
- Use as layers in Direct Tiles, OpenLayers, Leaflet, and Google Maps
- Available for Free and all other paid accounts

Weather stations

API doc (/api_station)

- Access recent data from more than 40,000 weather stations around the world

Figure 104: OpenWeatherMap home page.

7. Real-time weather forecasts: The easiest, most advanced, weather API on the web. (see Figure 105)

<https://developer.forecast.io/>



[The Dark Sky Forecast API](#)

- [API Docs](#)
- [Register](#)
- [Log in](#)

The easiest, most advanced, weather API on the web

The same API that powers [Forecast.io](#) and [Dark Sky for iOS](#) can provide accurate shortterm and longterm weather predictions to your business, application, or crazy idea.

We're developers too, and we like playing with new APIs, so we want you to be able to try ours hassle-free: all you need is an email address.

The API uses a [simple, JSON interface](#). Community-provided API wrappers enable you to integrate with just a couple lines of code!

Our simple pricing policy:

You can use the API in both commercial and non-commercial applications.

1. The first thousand API calls you make every day are free, period.
2. Every API call after that costs \$0.0001 each.
3. Credit us with a "Powered by Forecast" badge that links to <http://forecast.io/> wherever you display data from the API.



©2012-6 The Dark Sky Company, LLC
[Email Us](#) [Terms of Use](#) [Privacy Policy](#)

Figure 105: Real-time weather forecasts home page.

8. WeatherBug: With our APIs you can integrate the power of hyper localized data and the most comprehensive weather information into your products, services and apps. PulseAPI goes beyond typical weather data by providing the most reliable weather intelligence trusted by more than 40 million customers around the globe. We offer a wide range of options based on your volume and specific needs. Extend your product capabilities into a new dimension of intelligent weather offering with PulseAPIs. (see Figure 106)
<http://business.weatherbug.com/products/api-data-feeds/>

Weather (<http://weather.weatherbug.com>) • Earth Networks
 Sign In (<https://login.enterprise.weatherbug.com/>) Feedback (<mailto:customerfeedback@weatherbug.com>)
 WeatherBug (<http://business.weatherbug.com/>)
 Networks (<http://business.weatherbug.com/ournetworks/>) Products (<http://business.weatherbug.com/products/>)
 Solutions (<http://business.weatherbug.com/solutions/>) Events (<http://business.weatherbug.com/events/>)
 Support (<http://business.weatherbug.com/support/>)
 Industry's Most Extensive, Hyper-local, Weather Data Feeds

Extend Your Product Capabilities With Intelligent Weather Data

Our PulseAPI delivers real-time weather and lightning information, forecasts utilizing forecast models with the lowest margin of error, and other top-notch weather data. By partnering with us, you'll have the flexibility to use the data in your desired configuration. Data requests are made to a REST API, enabling quick and easy integration, and you can choose from a wide range of options based on your volume and specific needs.

As calls are made into the API, a wide range of response values are available with data returned via multiple formats. Each API Client is provided a token for secure access to their data.

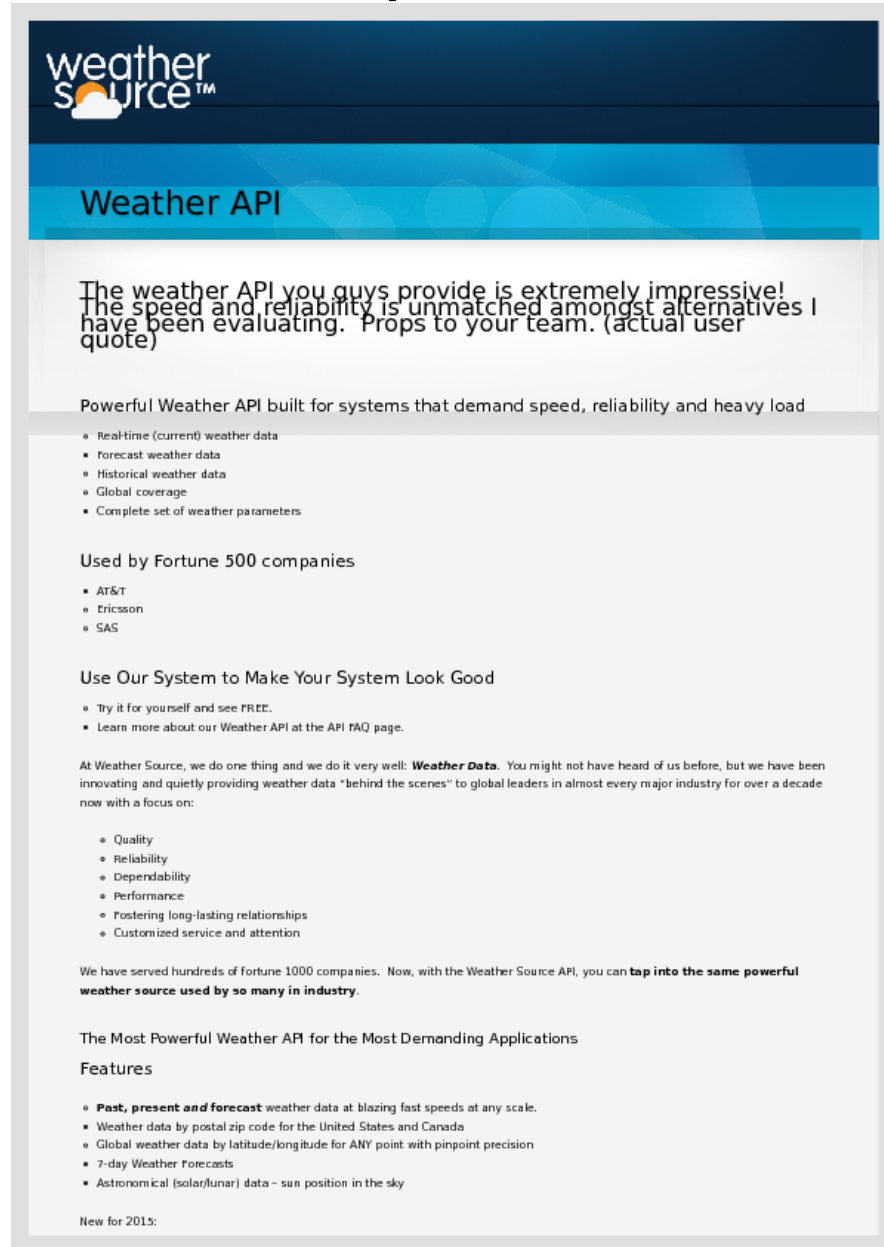
Customize your apps, websites and more with our PulseAPIs

Basic	Plus	Pro
Threshold: 150,000 / mo Transactions: 5 / sec \$20 /mo	Threshold: 150,000 / mo Transactions: 5 / sec \$150 /mo	Threshold: 150,000 / mo Transactions: 5 / sec \$300 /mo
10 Day-Night Forecast Location Search Real Time Weather Observations Search Weather Station Sky Conditions Icons Station List US Alerts (National Weather Service)	Basic and Air Quality Forecast (AQI) Hourly 6-Day Forecast Radar and Maps Sunrise and Sunset Times Ultraviolet Index (UV)	Basic and Plus and Bing Map Layers Google Map Layers

Figure 106: WeatherBug home page.

9. Weather Source: Powerful Weather API built systems that demand speed, reliability, and heavy load. (see Figure 107)

<http://weathersource.com/weather-api>



weather source™

Weather API

The weather API you guys provide is extremely impressive! The speed and reliability is unmatched amongst alternatives I have been evaluating. Props to your team. (actual user quote)

Powerful Weather API built for systems that demand speed, reliability and heavy load

- Real-time (current) weather data
- Forecast weather data
- Historical weather data
- Global coverage
- Complete set of weather parameters

Used by Fortune 500 companies

- AT&T
- Ericsson
- SAS

Use Our System to Make Your System Look Good

- Try it for yourself and see FREE.
- Learn more about our Weather API at the API FAQ page.

At Weather Source, we do one thing and we do it very well: **Weather Data**. You might not have heard of us before, but we have been innovating and quietly providing weather data "behind the scenes" to global leaders in almost every major industry for over a decade now with a focus on:

- Quality
- Reliability
- Dependability
- Performance
- Fostering long-lasting relationships
- Customized service and attention

We have served hundreds of fortune 1000 companies. Now, with the Weather Source API, you can **tap into the same powerful weather source used by so many in industry.**

The Most Powerful Weather API for the Most Demanding Applications

Features

- **Past, present and forecast** weather data at blazing fast speeds at any scale.
- Weather data by postal zip code for the United States and Canada
- Global weather data by latitude/longitude for ANY point with pinpoint precision
- 7-day Weather Forecasts
- Astronomical (solar/lunar) data - sun position in the sky

New for 2013:

Figure 107: Weather Source home page.

10. Weather Underground: Reliable data, accurate forecast, & global coverage in 80 languages.
(see Figure 108)

<http://www.wunderground.com/weather/api>

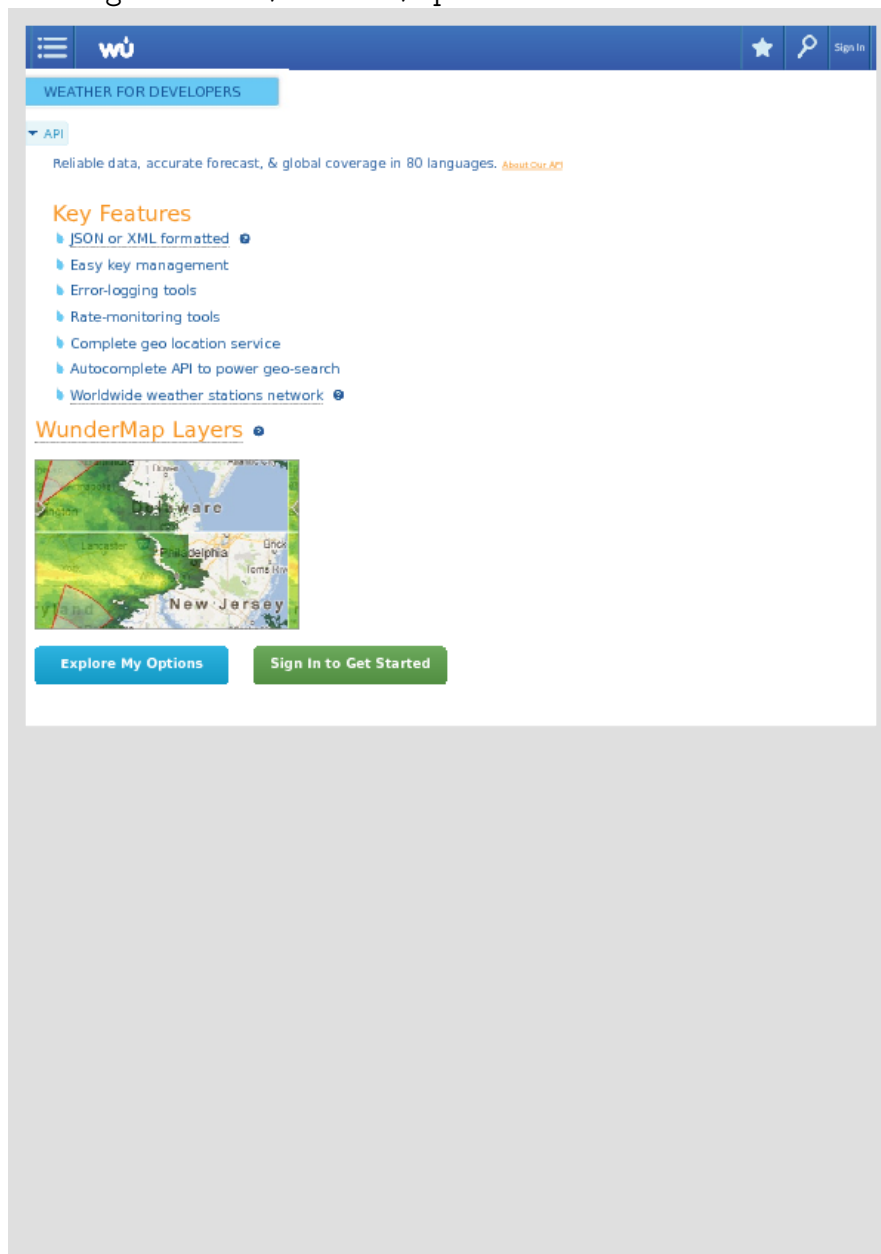



Figure 108: Weather Underground home page.

11. World Weather Online: World Weather Online's weather API (application programming interface) allows developers and programmers to access current, past and future weather data for use in apps and on websites. (see Figure 109)
<https://developer.worldweatheronline.com/>

Wishing everyone a joyful and blessed Eid-ul-Fitr.


 [Home](#) [API](#) [About Us](#) [Contact Us](#)


PREMIUM WEATHER API


TRY FREE FOR 60 DAYS
[\(/SIGNUP.ASPX\)](/SIGNUP.ASPX)

PRICING
[\(/API/PRICING.ASPX\)](/API/PRICING.ASPX)

JOIN NOW [\(/SIGNUP.ASPX\)](/SIGNUP.ASPX)

 **150,000+ API USERS**

 **BILLIONS OF WEATHER REQUESTS PER DAY**



Worldwide City and Town Weather API [read more \(/api/local-city-town-weather-api.aspx\)](/api/local-city-town-weather-api.aspx)
Current weather plus up to 14 day weather
1, 3, 6, 12 and 24 hourly interval
Monthly climate average data

Past/Historical Weather API [read more \(/api/historical-weather-api.aspx\)](/api/historical-weather-api.aspx)
Data from 1st July 2008, hourly interval

Figure 109: World Weather Online home page.

12. World Weather OnLine: Our weather API allows developers and programmers to embed weather data into their applications and websites. The local weather API provides access to present weather conditions and up to 15 days weather forecast across worldwide locations. (see Figure 110)

<http://us.worldweatheronline.com/api/local-city-town-weather-api.aspx>

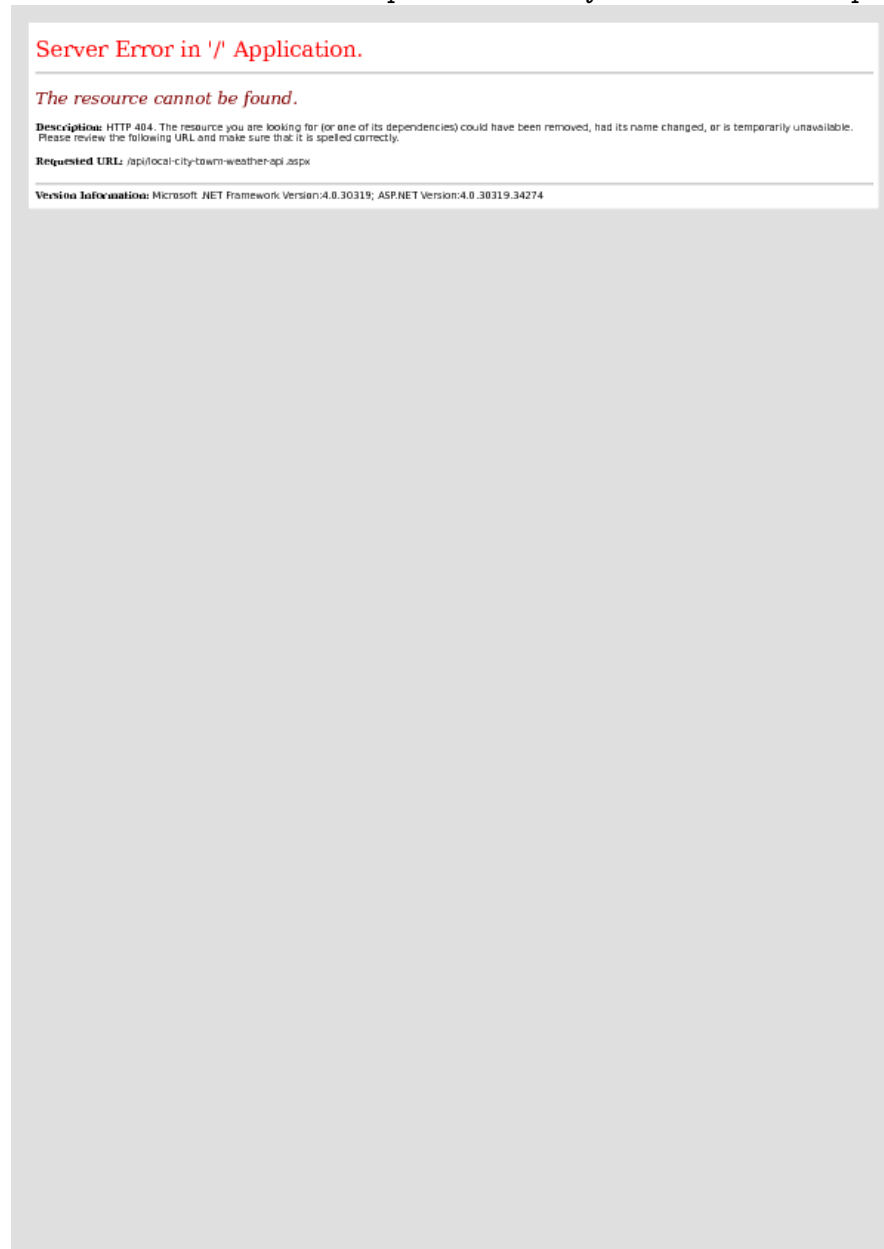


Figure 110: World Weather OnLine home page.

3.11 Zip code

1. Zip Codes: Our ZIP Code Database can be downloaded in both Microsoft Excel and CSV formats and easily opened in most spreadsheet applications or imported into the database software of your choosing. The data comes from authoritative sources such as the United States Postal Service (2011), US Census Bureau (2010), the Internal Revenue Service (2008), and Yahoo. (see Figure 111)

<http://www.unitedstateszipcodes.org/>

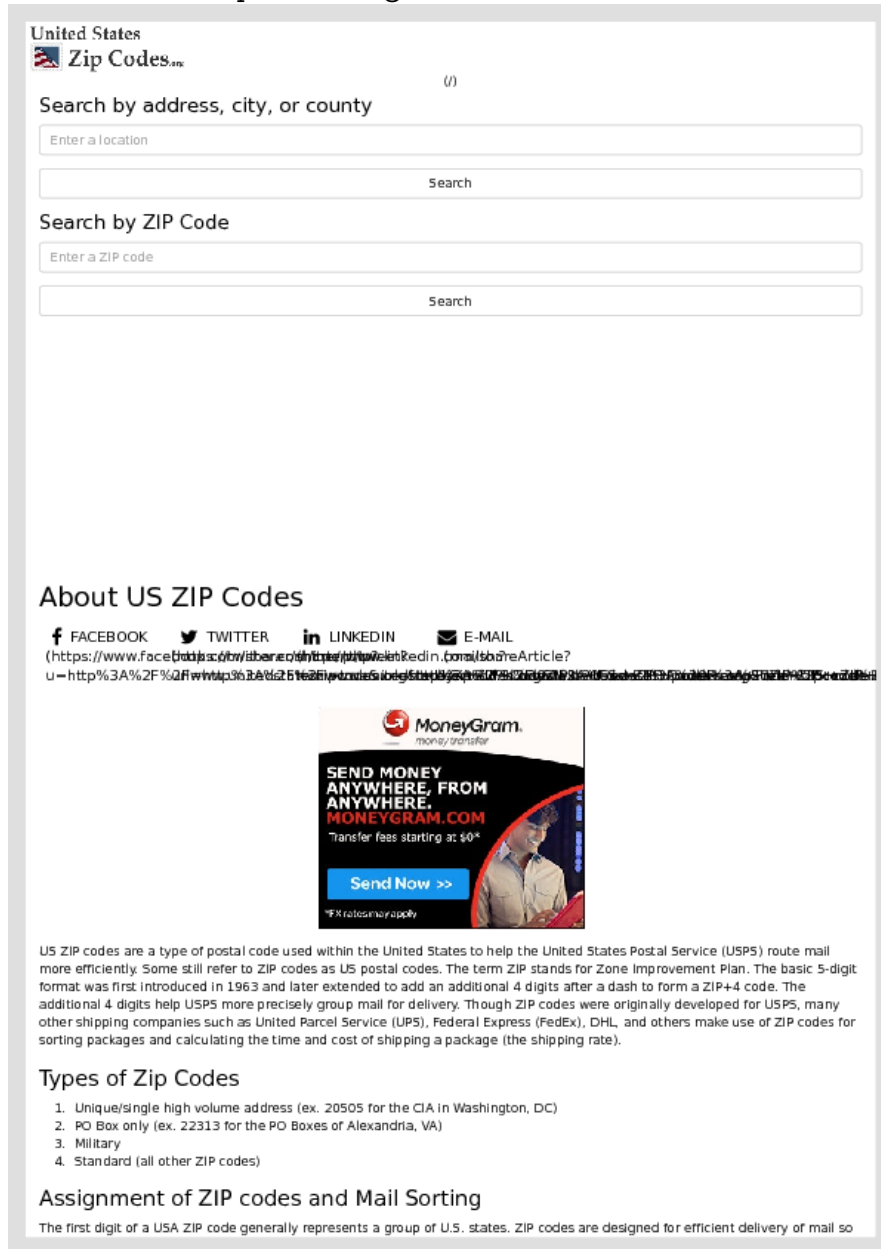


Figure 111: Zip Codes home page.

4 System performance

A graph showing system performance creating the home page images for all BD sources in this report has been created (see Figure 112).

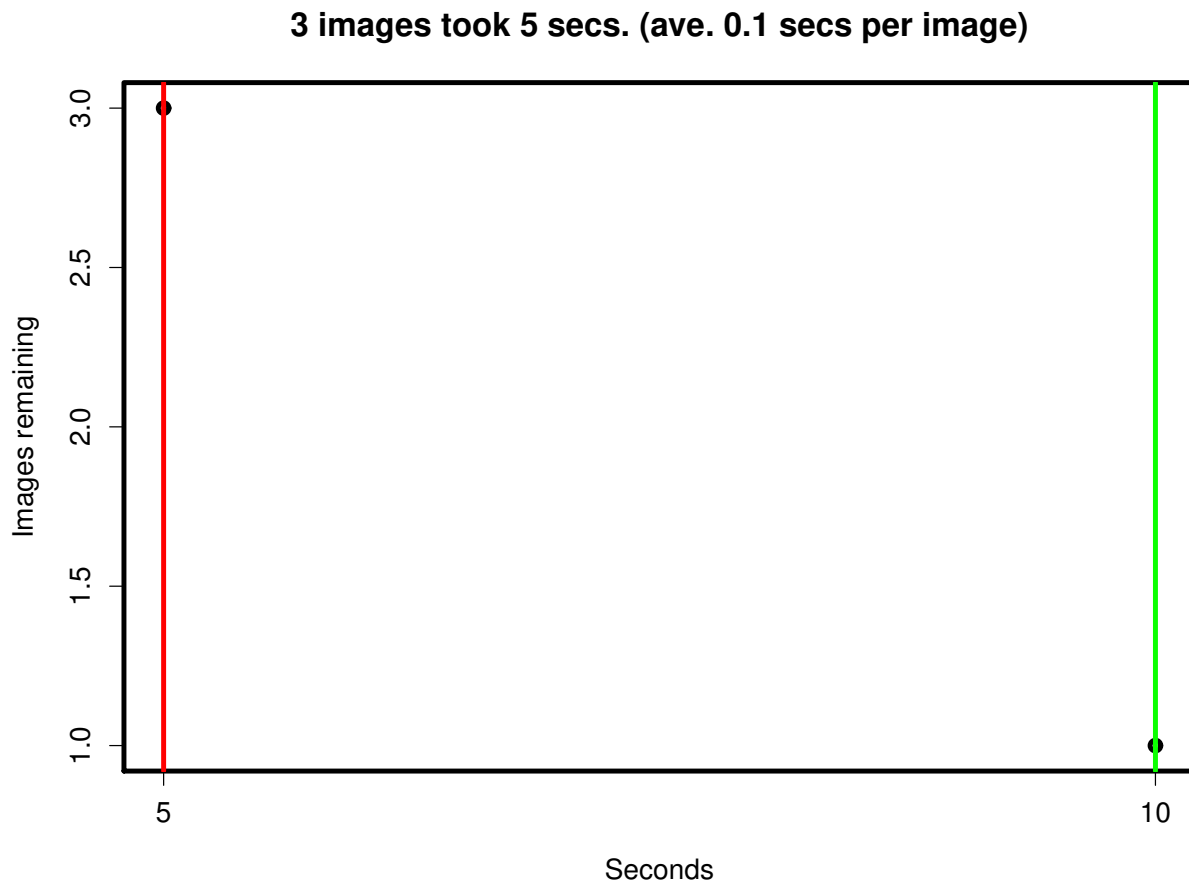


Figure 112: How long it took to create the home page images. The red lines are when 50 percent of the images were created. The green vertical line is when the first image was created.

5 References

- [1] Nicholas Felton, <http://feltron.com/>, 2014.
- [2] Viktor Mayer-Schönberger and Kenneth Cukier, *Big data: A revolution that will transform how we live, work, and think*, Houghton Mifflin Harcourt, 2013.
- [3] Aaron Parecki, <http://aaronparecki.com/>, 2015.