

The Data Collective makes public data trivially easy to find, understand, share and discuss. Anybody can post a data set, rate it, comment on it, and produce a graph to share in any medium - tweet, status update, blog post or article. Our goal is to make public discourse better informed by the facts.

By radically lowering the barriers to using real data in public discourse, we can change the competitive dynamics of blogging and journalism to favor those who back up their claims.

Early phases of the data would be seeded by the core team, with automation; ultimately, we would promote curation from a volunteer community. We would be a donor-backed 501(c)3, with ambitious goals for uptake by bloggers, journalists and citizens globally.

Sites we love:

Timetric.com -- Everything you need to publish data and research online. See what we do right now. We've loaded more than three million [public statistics](#) into Timetric. Great support for integrating a chart into online media. As of 20110602, they confirmed that they no longer support data uploading.

Wolfram Alpha. [Long-term goal](#) is to make all systematic knowledge immediately computable and accessible to everyone. Accepts data, but no self-service; all contributions are vetted. No online media integration.

Google [Public Data Explorer](#) makes large datasets easy to explore, visualize and communicate... [enabling] people to better understand the world around them and make more informed, data-driven decisions. They have proposed DSPL, [Dataset Publishing Language](#). Accepts uploads. Beta started Feb 16, 2011.

ManyEyes: [Our goal](#) is still to “democratize” visualization and to enable a new social kind of data analysis. NYTimes Economix [uses](#) ManyEyes.

GapMinder's [goal](#) is to replace devastating myths with a fact-based world view. 438 data sets as of Feb 16, 2011. No public posting.

Multpl.com is a brilliantly simple solo effort with 16 high-level economic datasets such as historical P/E ratio for US equities. No public posting. Solo side-project effort by Josh Staiger, former IBM & Google, now co-founder of grocery deal shopping site AnyLeaf.com.

zFacts hopes to fill a crucial but almost empty niche: To provide think-tank quality information in an easily accessible format. Our intention is to help the interested public and the press understand current social and political issues.

BuzzData (in private beta as of 20110610) is a web platform in which data, issues and people converge in an intuitive, streamlined manner, such that innovation is rapid and collaboration is easy. We're putting data front and centre, where it should be.

Datacurious (in development) Controversial trends. Visualized. Compared. With Nuance. [DJ says: currently focused on data about private companies to help investors]

[iCharts](#) is a software-as-a-service (SaaS) company that provides the only easy-to-use and affordable online service for interactive chart creation and social sharing. Free service launched June 2011. Requires login & email verification. Flash. "launched as a TechCrunch50 Finalist in Fall 2008. iCharts vision is to empower the world to explore and share data by enabling the free flow of visual data among data owners, media, business and consumers."

[Explainum](#) is a web-service for creating live embeddable charts with comments. Unlike many charting tools Explainum gives you the ability to annotate data points which is useful for collecting and sharing knowledge about specific trends. It's like twitter or wiki for charts. Sign-in required. Daily data only. 3 person startup.

[Wondergraphs](#) strives to be [the best way to get and share insights from your data](#). Tried and failed to upload some data.

Verifiable.com no longer operating. Beautiful UI. Post-mortem [here](#).

Swivel.com no longer operating. Venture-backed. Halsey Minor. Post-mortem [here](#).

Economy app is ipad-only, currently. Beautiful and effective UI. #1 Finance App for 6 weeks, [#12 on 3/8/11](#). Data presentation has problems; for example, some long-term data is presented with a linear y-axis, so all you can see is the fact that there was exponential growth.

Numbrary.com is a free online service dedicated to finding, using and sharing numbers on the web. Very clean UI and concept, uses amCharts. Blogging stopped in 2008. Founder started new venture in Jan 2010. Posting data works but it is not automatically charted as of 201106. DJ emailed them 201106 about this.

Data360.org is free and features about 5,000 relevant datasets, with graphing features. Comments are overwhelmingly spam. DJ finds it hard to use.

NationMaster.com and **StateMaster.com** offer data on nations and US states.

WikiPosit [aims](#) to be a "Wikipedia" of numerical data. DJ finds it hard to use, but many of its ideas are solid. Solo side-project effort by Tammer Kamel, an investment consultant.

Factual.com is an open data platform and community in an effort to maximize data accuracy, transparency, and availability. Raised \$25M Dec 2010 from Andreessen Horowitz and Index Ventures. Ron Conway et al.

[GetTheData.org](#). Ask and answer data related questions.

The [Socrata Social Data Platform](#)™ is the most widely-adopted Open Data solution in Government. Socrata customers include Medicare, State of Washington and City of Seattle. Socrata was recently chosen to deliver the next-generation Open Data platform for [Data.Gov](#) and participating federal agencies, through its GSA reseller and fulfillment partner, Alamo City Engineering Services.

<http://open.mflask.com/> is a free search engine for data sets published by governments, private companies and other organizations. It now indexes [255473 datasets](#) from many sources.

[ScraperWiki](#) is all the tools you need for Screen Scraping, Data Mining & visualisation. Make bad data good, collaborate & discover new datasets.

[Windows Azure DataMarket](#): **One stop shop for Data**. Get all data you need for your insights: trusted commercial and premium public domain data. End-user data access requires Windows Live ID and either Excel 2010, Tableau, or any of several Microsoft BI tools. Developers can access data using Visual Studio or any development tool that supports HTTP.

InfoChimps: **We make data easy to access so you can focus on being more productive**. Query data sets half a terabyte in size without ever needing Cassandra, HBase, or even MySQL.

Tableau Public runs only on Windows. Can someone play around with it and describe it briefly here?

d8taplex
ckan

Scrapyard of verbiage from earlier descriptions of The Data Collective:

The Data Collective aims to radically simplify the discovery and use of data in everyday communication. ... will supply the graphs and data journalists need to tell the world's stories. By radically lowering the barriers to using real data in public discourse, we can change the competitive dynamics of blogging and journalism to favor those who back up their claims. Early phases of the data would be seeded by the core team, with automation; ultimately, we would promote curation from a volunteer community of subject matter experts, (ie grad students). We would be a donor-backed 501(c)3, with ambitious goals for uptake by bloggers, journalists and citizens globally.

data.org aims to elevate substantially the quality and presentation of data available to the journalistic universe. The idea is to create a central repository where anybody can access a wide variety of data, presented in an easy-to-use and easy-to-interpret format.

Scrapyard of questions about any site purporting to offer similar services:

Can anybody post?

Is it easy to post a dataset?

When the source data changes/vanishes, do reasonable things happen?

Is it easy to comment/vote on a dataset?

What incentives does someone have to post?

Can users search?

Are the results relevant?

Is the process smooth & easy & logical?

Can users browse?

Are the categories laid out logically?

Is the process smooth & easy & logical?

Can users get a graph to use in email, web, blog, print?

Do the graphs look good?

Does the graph link back to the source data?

Is the graph interactive or link to interactive?

Is the process smooth & easy & logical?

Can the source data opt out?

Sparklines

Linked numbers

Voting/commenting on a dataset

Voting/commenting on a graph

Achievements/recognition for everything