



Open Data Platform Evaluation and Selection

City and County of Durham's Open Data Program

BY JASON HARE: OPEN DATA PROGRAM ADVISOR



Table of Contents

Executive Summary	4
Reactive versus Proactive Data as a Service.....	4
Purpose and Methodology	5
Outcomes of the Evaluation	5
Expert Consultations used to create this section:	5
Establish a SaaS open data platform that is optimized to enable the reuse and dissemination of open data through non proprietary formats.....	6
Benefits of Cloud Computing.....	6
Types of Cloud Services	7
Standards and Guidance	7
Technology for Open Data	7
Requirements and evaluation	7
Open Data Portal Solutions	8
Methodology	8
Statement of work Functional requirements	9
Evaluation by Vendor.....	11
Socrata	11
Socrata's Commitment to Interoperability	12
Integrate and federate CKAN catalogs with the Socrata platform	12
Socrata's evaluation on requirements.....	12
CKAN	13
CKAN's Evaluation on requirements.....	13
Open Data Soft	14
Open Data Soft's Evaluation on requirements.....	14
Recommendations	16
Technology for Open Data Tools - Tools Evaluation.....	16
Socrata Scores	17
CKAN Scores.....	21



Open Data Soft Scores.....	24
Open Data Definition of Key Terms/Acronyms.....	27



Executive Summary

Open Data is the idea that data should be freely available for everyone to access, use and republish as they wish, published without restrictions from copyright, patents or other mechanisms of control. Public sector information made available to the public as open data is termed 'Open Government Data'.

Governments and their contractors collect a vast quantity of high-quality data as part of their ordinary working activities. Typically this results in the jurisdiction(s) becoming a powerful data monopoly able to structure and homogenize the interactions between itself and its citizens. These one-sided interactions are expensive and unresponsive to citizens' needs and can unnecessarily restrict government activities, as well.

Opening government data involves both policy and technical considerations. If governments' data are made open, they can have huge potential benefits including:

- **Transparency:** In a well-functioning, local government citizens need to know what their government is doing. To do that, they must be able freely to access government data and information and to analyze and share that information with other citizens.
- **Efficiency:** Enabling better coordination and efficiency within government, by making data easier to find, analyze and combine across different government units and agencies.
- **Innovation:** In a digital age, data is a key resource for social and commercial activities. Everything from catching a bus to finding a doctor depends on access to information, much of which is created or held by government. By opening up data, government can help drive the creation of innovative business and services that deliver social and commercial value.

More about definitions can be found at www.opengovguide.com/glossary

Reactive versus Proactive Data as a Service

Where many public records laws and policies regulating the right to information have traditionally relied on reactive disclosure, meaning public information has to be requested before it is shared, a government fully engaged in open data is choosing to proactively disclose information - meaning public data is released as it is collected and before it is requested. Put another way, the vision of open data is for government information to be 'open by default'. Open data also have a number of technical implications, with special consideration given to the particular formats chosen for data release. Open



formats are those that are structured and non-proprietary, allowing the public and the government to extract maximum value from the information now and in the future.

Governments around the world and across the US cite many different reasons for starting open data initiatives, including increasing government transparency and accountability, catalyzing the creation of new digital services and applications for citizens, unlocking the full economic potential of public information, and evolving current government services for anticipated future needs. Although much of this top-level government interest is new, there are many professions and communities engaged in dialogue, policy, and development around this issue, including from government officials, journalists, developers, transparency reformers, issue advocates, and interested citizens.

Purpose and Methodology

There are several dozen versions of open data platforms and API architectures that can be used to build an effective open data program. This report has selected and evaluated three viable options based upon the criteria established by the City and the County CIOs. The three solutions are CKAN, Socrata and Open Data Soft. The methodology for evaluation was a self-evaluation by the vendors and a subsequent evaluation by the author of this paper. The evaluation rubric includes score weighting..

Outcomes of the Evaluation

- OpenDataSoft scored the highest with 1653
- CKAN scored 1439
- Socrata scored 1199

The recommendation is OpenDataSoft hosted on Azure Dublin. Azure Dublin is the largest single cloud instance to date. OpenDataSoft has successfully deployed its platform on Azure.

Expert Consultations used to create this section:

- The Open Data Institute: An open data think tank
<http://theodinc.org>
- Global Open Data Initiative: A consortium of several open data organizations
<http://globalopendatainitiative.org/>
- Sunlight Foundation: A transparency organization that provides support to the OGP
<http://sunlightfoundation.com/>
- Open Knowledge Foundation: A open data thought leader and the parent company of CKAN
<http://okfn.org/>



- World Bank's Open Government Data Working Group: Working on open data projects around the world, the OGDWG has much field experience and shares it through online resources <http://data.worldbank.org/>
- Web Foundation: A member of the Global Open Data initiative <http://www.webfoundation.org/>
- Open Institute: A member of the Global Open Data Initiative <http://openinstitute.com/>
- OGP Working Group: Open Data: The author is a member of the Open Data Working Group for the Open Government Partnership. This working group concerns itself with the emerging global standard that is curated by the Open Data Institute. <http://www.opengovpartnership.org/get-involved/join-working-group>

Establish a SaaS open data platform that is optimized to enable the reuse and dissemination of open data through non proprietary formats

On premise or off-premise Software as a Service is generally preferred to host open data portals. SaaS is often referred to as cloud computing.

Simply put, cloud computing is the industry term for delivering hosted services over a network or the Internet. It treats computing as a service rather than a product, enabling users to access and share a wide variety of applications, data, and resources through an interface such as their web browser. Cloud computing requires a paradigm shift in infrastructure management. Instead of local PCs running individual applications, cloud IT services tap a distributed network of computers that may reside in multiple locations yet appear as a single resource on the front end. In cloud infrastructure solutions, everything becomes a software service – applications, data, and storage – and APIs are the building blocks. Obviously, security on the cloud is critical for both businesses and users.

Benefits of Cloud Computing

The goal of cloud computing is to provide easy, scalable access to computer resources and services. The potential benefits for enterprise adoption of cloud services are huge and include:

- Decreased infrastructure costs
- Reduced time to market
- Flexibility in infrastructure investments beyond today's virtualization solutions
- Ability to rapidly adopt and apply game-changing technologies
- Enhanced partnering opportunities due to increased business interoperability
- Full compatibility with the trend away from PCs to portable and purpose-specific devices

City of Durham | 101 City Hall Plaza, Durham, NC
27701
(919) 560-1200 | durhamnc.gov

Durham County Administrative Complex | 200 East
Main Street, Durham, North Carolina 27701
(919) 560-7015 | Fax (919) 560-7078 | dconc.gov



Types of Cloud Services

Cloud IT services are usually divided into three categories:

- Infrastructure as a Service (IaaS). These cloud services offer configurable virtual servers and storage, and companies or individual users pay for the capacity they use, which can fluctuate as needed. Example: Amazon Web Services.
- Platform as a Service (PaaS). The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.
- Software as a Service (SaaS). Here, the host provides everything – infrastructure, software, data storage, and the user interface. Because everything is in the cloud, users can access the services from anywhere. Examples: Web-based email services such as Socrata, Open Data Soft and CKAN all provide turn-key web service solutions.

Standards and Guidance

Throughout this document the activities and recommendations are based on the OGP set of standards and practices. In particular the following experts were consulted:

- Open Knowledge Foundation Open Data Handbook
<http://opendatahandbook.org/pdf/OpenDataHandbook.pdf>
- Sunlight Foundation: Open Data Policy Guidelines
<http://sunlightfoundation.com/opendataguidelines/>
- World Bank Open Data Toolkit
<http://data.worldbank.org/open-government-data-toolkit>
- TOGAF™ -- The Open Group Architecture Framework
<http://pubs.opengroup.org/architecture/togaf8-doc/arch/>

Technology for Open Data

Requirements and evaluation

The requirements for the open data portal platform were communicated with me. These are summarized below and following the criteria is a description of each product used in the evaluation.



Open Data Portal Solutions

- SOCRATA; SaaS based data visualization platform based in Seattle, USA. This is a proprietary solution.
- CKAN; On-premise or SaaS based data catalog platform based in London, England. This is an open source solution.
- Open Data Soft; On-premise or SaaS based data management platform with catalog, API and visualization tools based in Paris, France. This is an open-source solution.

Methodology

The methodology will be a weighted matrix written in Excel. Each data platform will be evaluated using the same criteria. The weight of each facet of the data platform can be adjusted once the assessment team has studied the list of requirements. Not any one platform will have all of the functionality listed in the requirements out of the box. Each platform vendor will be given the chance to comment on the evaluation criteria. The assessment team will approach each vendor and keep the client anonymous in regards to questioning the vendor(s).



Statement of work Functional requirements

Querying and discovery	Data Application and Management	Performance Management and Maintenance	Social Engagement and Community Functional Requirements
<ul style="list-style-type: none"> • Open data query • Application query • SPARQL query • Content management • Events calendar • Comments and data voting • Comments and application voting • Notifications • Blogs and forums • Changing of Data format • Application integration • Social media integration 	<ul style="list-style-type: none"> • Data load & publishing (inventory) • Data load & publishing (datasets) • Recording & application approval • Data & application download 	<ul style="list-style-type: none"> • Security Audits • Error control • Administration • Tracking & monitoring 	<ul style="list-style-type: none"> • A specialized site for developers • Available open API based on standards • Community for developers • Community forums • Code and application examples • Blogs for developers • Load & query on the app store • Metadata management for datasets • Format transformations, geocoding



			<ul style="list-style-type: none">● Cloud storage and processing● Portal content management● Site customization● Role management● Data federation services (to share data by exposing and/or by metadata catalogs)● Shared resources● Report generation● Dataset digitally signed● Multiple data level routing and approval workflows to publish● Dataset download at least as CSV, JSON, PDF, RDF, RSS, XLS, XLSX, or XML formats● Map
--	--	--	---



			<p>visualization (Google, Bing, ESRI)</p> <ul style="list-style-type: none">● Graph and plots (area, bars, columns, doughnut, lines, circular, timeline, tree map, bubbles)● Converters and plugins● Citizenship participation● Linking, tagging data● RDF graphs
--	--	--	---

Evaluation by Vendor

Socrata

Socrata has been supporting the open data movement since 2008. The Socrata platform effectively opens up data to residents, developers, the media, academia, and advocacy groups, by providing support for these open data standards:

- Open Discovery: anyone can easily find the data they're interested using robust search, catalog browsing and search engine optimization to increase the discoverability of data.
- Open Formats and Protocols: any dataset on the Socrata platform is instantly and automatically downloadable (through the user interface or the API) in a number of open file formats including: CSV, JSON, RDF and XML. That's in addition to XLS/XLSX, RSS, PDF and other commonly used formats.
- Metadata Standards: Socrata offers an open, robust, and configurable metadata management system that allows our customers to conform to the catalog and data standards they wish to follow, such as the W3C DCAT standard.

City of Durham | 101 City Hall Plaza, Durham, NC 27701
(919) 560-1200 | durhamnc.gov

Durham County Administrative Complex | 200 East Main Street, Durham, North Carolina 27701
(919) 560-7015 | Fax (919) 560-7078 | dconc.gov



- Open APIs: Every dataset in Socrata is automatically accessible using an open, standards-based and non-proprietary API, called the Socrata Open Data API, or SODA. Since 2010, the SODA specifications were released to the Open Data community under a Creative Commons license. As a result, SODA is now the most widely used open data API in the world, with hundreds of applications making millions of API requests every month.

Socrata's Commitment to Interoperability

This author believes that the success of the open data movement requires data portability throughout the ecosystem. To that effect, Socrata has invested in several initiatives and technologies that facilitate the flow of data between Socrata and other open data systems and catalogues.

Integrate and federate CKAN catalogs with the Socrata platform

With Socrata the ODP can bring any catalogs together by using an open source tool called the Data-Importer. This simple, standalone tool periodically syncs a CKAN catalog into a Socrata instance. A great example of this can be found at: data.gov/cities/Community/Cities/Datasets. This site, powered by Socrata, integrates other Socrata-based government catalogues with catalogues based on the CKAN Project.

It is equally as easy to integrate Socrata-powered resources – datasets, maps, charts, APIs, dashboards, and other data-sharing apps – into catalogue management tools like the CKAN Project. To simplify this process, Socrata provides:

- Support for standards such as DCAT and Project Open Data standards. These standards already have harvesters available. Learn more at: github.com/HHS/ckanext-datajson
- A direct Socrata harvester for CKAN sites. Learn more at: github.com/socrata/socrata-harvester

Aggregation catalogs like data.gov and healthdata.gov provide great examples of this kind of Socrata-to-CKAN integration.

Socrata's evaluation on requirements

Socrata was evaluated with the following results, based on the requirements specific scores, which are referenced in in the evaluation piece of this document:

- **Querying and Discovery & Data Application Management:** Socrata has all the basic functionality for aggregating and querying data, publishing and download data and applications.
- **Platform Management and Maintenance:** Socrata has a very elegant admin model that allows a very good error control and administration of datasets and organizations.



- **Social Engagement & Community:** Socrata has invested in its features for Social Engagement and Community; the only functionality that does not meet for the moment is the creation of RDF graphs.
- **Support:** Socrata has been on the market for a good time and provides a good amount of support for its customers.
- **Total Score:** 1199

CKAN

CKAN is part of a larger set of tools and services designed to enable automated use and reuse of content and data:

The key entity in CKAN is a Package. The data model for a package is pretty much what you see on: <http://www.ckan.net/package/new>

This in turn is heavily based on the kind of packaging information provided for software but with some modifications. One of the design aims is to keep things simple and as generic as possible as data originates from a lot of different domains. Thus the product is designed to keep the core metadata pretty restricted but allow for additional info either via tags or via “extra” arbitrary key/value fields. So very roughly:

- unique name
- title
- url + download url
- author/maintainer info
- license
- notes
- tags
- [extendable with “extra” fields]

CKAN’s Evaluation on requirements

CKAN was evaluated with the following results:

Querying and Discovery & Data Application Management: CKAN has all the basic functionality for querying and discovery of data, however it lacks a pretty important functionality as changing format.

- **Platform Management and Maintenance:** The administration mode of CKAN needs a couple of customizations for having a good functionality for tracking and monitoring.



- **Social Engagement:** CKAN by default does not support a lot of social engagement and community out of the box. It does have good functionality for tagging linking data, creating graphs and plots. However, if the local government wants to improve citizenship participation it will need to invest in modifying CKAN theme to a friendlier one.
- **Support:** CKAN is an Open Source Platform. However, the OKFN offers programs to support CKAN for governments.
- **Total Score:** 1438

Open Data Soft

OpenDataSoft is a French startup, which develops a full-featured data-centric platform for helping its customers to enter straight in the cloud and big data era. Founded in 2011 by specialists of very large scale data processing, highly experienced in developing innovative services on the web (media, eCommerce, industry, governments), OpenDataSoft has a unique footprint to support its customers in implementing a real data strategy: internal and external data sourcing and sharing (API architectures, crowd sourcing), data transformation into business assets (interoperability, consolidation, data anonymizing,), quick development of innovative services & business models based on data valorization.

OpenDataSoft's main objective is to provide to its customers a turnkey solution to help them transform their internal data into valuable accessible assets.

Open Data Soft (ODS) is a SaaS proprietary solution that is a step further in regards to being a data platform rather than a data visualization and indexing platform. (See opendatasoft.com for a demonstration)

CKAN and Socrata are also SaaS enabled. CKAN is a true data repository capable of storing and sharing data through triple linked RDFs. Socrata can use pseudo RDF links but excels at the data visualization pieces. Where ***ODS excels is in data provenance and on the fly data unions and joins through its API cloud. Open Data Soft can power other data visualization platforms and comes with several hundred built in data connectors.***

ODS does have its own data visualization library but acknowledges some of the visualizations are highly resource intensive (like heat maps). It is cloud agnostic and can run on premise.

Open Data Soft's Evaluation on requirements

Open Data Soft was evaluated with the following results:



- **Querying and Discovery & Data Application Management:** Includes most of the functionality for querying and discovery but it lacks a couple of specialized pages, like one for developers and events calendar.
- **Platform Management and Maintenance:** Even though, it is a new platform Open Data Soft has implemented several features that support the administration of the portal.
- **Social Engagement:** Open Data Soft has implemented several nice functionalities out of the box for engagement; for example, they already implement Disqus for comments and voting in datasets. An extension that is widely used in the modern website. However, it lacks the integration of a CMS, although, they already implement several of the capabilities.
- **Support:** Open Data Soft offers a complete support for its platform.
- **Total Score:** 1653



Recommendations

Technology for the Platform

After the evaluation performed, I recommend Open Data Soft as the platform that should be used by to implement our Open Data Initiative. Open Data Soft proved to be a solid platform with all the features of the other portals evaluated and some additional attributes:

1. API First and based on search technology
2. Cost advantage
3. No discernable latency issues with large data sets
4. Fully customizable on the site design and work flow

The recommended OpenDataSoft solution is based on a SaaS cloud model hosted on the ODS Azure cloud. This project should consult with City and County's Cloud Security policy.

Technology for Open Data Tools - Tools Evaluation

Methodology of evaluation

Everything was weighted based on 1-5 based on:

5 - the functionality is out of the box.

4 - functionality is not out of the box, but with little development it can be included.

3 - functionality requires a middle effort to be added.

2 - functionality requires a big effort to be added.

1- functionality requires a major effort to be added.

0 - The functionality is not there and will require a big effort to be added with major complications.



Socrata Scores

The following are the scores that the team assigned to Socrata's Open Data Platform.

	Weighting (1-10)	Socrata (0-5)	Weighted Score for Socrata
Querying and Discovery			
Open Data Query (ODATA)	10	5	50
Application Query	10	5	50
SPARQL Query	3	1	3
Content Management	10	1	10
Events Calendar	5	2	25
Comments and data voting	5	5	25
Comments and application voting	2	5	10
Notifications	9	5	45
Blogs and forums	3	5	15
Changing of Data format	10	1	10
Application integration	8	2	16
Social media integration	10	2	20
Data & Application Management			



Data load & publishing (inventory)	10	5	50
Data load & publishing (datasets)	10	5	50
Recording & application approval	9	1	45
Data and application download	10	5	50
Subtotal			464
Platform Management and Mainten			
Security	10	5	50
Error Control	10	2	20
Administration	10	1	10
Tracking and Monitoring	10	1	10
Subtotal			90
Social Engagement and Community			
A specialized site for developers	8	5	40
Available open API based on standards	10	5	50
Community for developers	5	5	25
Community forums	5	5	25
Code and application examples	10	5	50
Blogs for developers	10	0	0
Load & Query on the app store	10	5	50
Metadata management for datasets	10	2	10

City of Durham | 101 City Hall Plaza, Durham, NC 27701
(919) 560-1200 | durhamnc.gov

Durham County Administrative Complex | 200 East Main Street, Durham, North Carolina 27701
(919) 560-7015 | Fax (919) 560-7078 | dconc.gov



Formats transformations, geocoding	10	2	20
Cloud storage and processing	10	2	20
Portal content management	10	3	30
Site customization	10	1	10
Role management	10	1	10
Data federation services (to share data by exposing and/or by metadata catalogs)	10	5	50
Shared resources	10	5	50
Report generation	8	0	0
Dataset digitally signed	2	0	0
Multiple data level routing and approval workflows to publish	5	1	5
Dataset download at least as CSV, JSON, PDF, RDF, RSS, XLS, XLSX or XML format	10	5	50
Map visualization (Google, Bing, ESRI)	10	5	50
Graph and plots	10	5	50
Converters and plugins	10	1	10
Citizenship participation	10	2	20
Linking, tagging data	10	2	20
RDF graphs	5	0	0
Total			1199



City of Durham | 101 City Hall Plaza, Durham, NC
27701 |
(919) 560-1200 | durhamnc.gov

Durham County Administrative Complex | 200 East
Main Street, Durham, North Carolina 27701
(919) 560-7015 | Fax (919) 560-7078 | dconc.gov



CKAN Scores

The following are the scores that the team assigned to CKAN's Open Data Platform.

	Weighting (1-10)	CKAN Score (0-5)	Weighted Score for CKAN
Querying and Discovery			
Open Data Query	10	5	50
Application Query	10	5	50
SPARQL Query	3	0	0
Content Management	10	1	10
Events Calendar	1	3	15
Comments and data voting	2	3	6
Comments and application voting	2	3	6
Notifications	9	3	27
Blogs and forums	3	2	6
Changing of Data format	8	0	0
Application integration	8	5	40
Social media integration	10	5	50
Data & Application Management			
Data load & publishing (inventory)	10	5	50
Data load & publishing (datasets)	10	5	50
Recording & application approval	9	0	0
Data and application download	10	0	0

City of Durham | 101 City Hall Plaza, Durham, NC 27701
(919) 560-1200 | durhamnc.gov

Durham County Administrative Complex | 200 East Main Street, Durham, North Carolina 27701
(919) 560-7015 | Fax (919) 560-7078 | dconc.gov



Subtotal			371
<i>Platform Management and Mainten</i>			
Security	10	5	50
Error Control	10	5	50
Administration	10	4	40
Tracking and Monitoring	10	4	40
Subtotal			180
<i>Social Engagement and Community</i>			
A specialized site for developers	8	3	24
Available open API based on standards	10	4	40
Community for developers	8	3	24
Community forums	8	3	24
Code and application examples	10	3	30
Blogs for developers	10	3	30
Load & Query on the app store	10	3	30
Metadata management for datasets	10	5	50
Formats transformations, geocoding	10	3	30
Cloud storage and processing	10	5	50
Portal content management	10	3	30
Site customization	10	3	30



Role management	10	5	50
Data federation services (to share data by exposing and/or by metadata catalogs)	10	5	50
Shared resources	10	4	40
Report generation	8	4	32
Dataset digitally signed	2	2	4
Multiple data level routing and approval workflows to publish	5	2	10
Dataset download at least as CSV, JSON, PDF, RDF, RSS, XLS, XLSX o XML format	10	5	50
Map visualization (Google, Bing, ESRI)	10	2	20
Graph and plots	10	4	40
Converters and plugins	10	5	50
Citenship participation	10	3	30
Linking, tagging data	10	5	50
RDF graphs	5	2	10
Total			1429



Open Data Soft Scores

The following are the scores that the team assigned to Open Data Soft's Open Data Platform.

	Weighting (1-10)	Open Data Soft Score (0-5)	Weighted Score for Open Data Soft
Querying and Discovery			
Open Data Query	10	5	50
Application Query	10	5	50
SPARQL Query	3	0	0
Content Management	10	5	50
Events Calendar	5	0	0
Comments and data voting	10	5	50
Comments and application voting	2	5	10
Notifications	9	5	45
Blogs and forums	3	2	6
Changing of Data format	10	5	50
Application integration	10	5	50
Social media integration	10	5	50
Data & Application Management			
Data load & publishing (inventory)	10	5	50
Data load & publishing (datasets)	10	5	50
Recording & application approval	9	5	45



Data and application download	10	4	40
Subtotal			546
<i>Platform Management and Mainten</i>			
Security	10	5	50
Error Control	10	5	50
Administration	10	5	50
Tracking and Monitoring	10	5	50
Subtotal			200
<i>Social Engagement and Community</i>			
A specialized site for developers	8	4	32
Available open API based on standards	10	5	50
Community for developers	8	0	0
Community forums	8	0	0
Code and application examples	10	4	40
Blogs for developers	10	0	0
Load & Query on the app store	10	0	0
Metadata management for datasets	10	5	50
Formats transformations, geocoding	10	5	50
Cloud storage and processing	10	5	50
Portal content management	10	3	30



Site customization	10	5	50
Role management	10	5	50
Data federation services (to share data by exposing and/or by metadata catalogs)	10	5	50
Shared resources	10	5	50
Report generation	8	5	40
Dataset digitally signed	2	0	0
Multiple data level routing and approval workflows to publish	10	5	50
Dataset download at least as CSV, JSON, PDF, RDF, RSS, XLS, XLSX o XML format	10	5	50
Map visualization (Google, Bing, ESRI)	10	5	50
Graph and plots	10	5	50
Converters and plugins	10	5	50
Citenship participation	10	5	50
Linking, tagging data	10	5	50
RDF graphs	5	3	15
Subtotal			907
TOTAL			1653



Open Data Definition of Key Terms/Acronyms

This subsection provides the definitions of all terms, acronyms, and abbreviations referenced throughout this document.

- **API:** An application-programming interface, which is a set of definitions of the ways one piece of computer software communicates with another. It is a method of achieving abstraction, usually (but not necessarily) between higher-level and lower-level software.
- **API Analytics:** Rate limiting will be part of any API platform, without some sort of usage log and analytics showing developers where they stand, the rate limits will cause nothing but frustration. Clearly show developers where they are at with daily, weekly or monthly API usage and provide proper relief valves allowing them to scale their usage properly.
- **API Documentation:** Quality API documentation is the gateway to a successful API. API documentation needs to be complete, yet simple—a very difficult balance to achieve. This balance takes work and will take the work of more than one individual on an API development team to make happen. Developers of the API can write API documentation, but developers who were not responsible for deploying the API should make additional edits. As a developer, it's easy to overlook parameters and other details that developers have made assumptions about.
- **Application Library:** Complete, functioning applications built on an API are the end goal of any API owner. Make sure and showcase all applications that are built on an API using an application showcase or directory. App showcases are a great way to showcase not just applications built by the API owner, but also showcase the successful integrations of ecosystem partners and individual developers.
- **Basic Auth:** Basic Auth is a way for a web browser or application to provide credentials in the form of a username and password. Because Basic Auth is integrated into HTTP protocol it is the easiest way for users to authenticate with a RESTful API. Basic Auth is easily integrated, however if SSL is not used, the username and password are passed in plain text and can be easily intercepted on the open Internet.
- **Catalog:** A catalog is a collection of datasets or web services.



- **Code Library:** Working code samples in the entire top programming languages are commonplace in the most successful APIs. Documentation will describe in a general way, how to use an API, but code samples will speak in the specific language of developers.
- **Content API:** A web service that provides dynamic access to the page content of a website, includes the title, body, and body elements of individual pages. Such an API often but not always functions atop a Content Management System.
- **CSV:** A comma separated values (CSV) file is a computer data file used for implementing the tried and true organizational tool, the Comma Separated List. The CSV file is used for the digital storage of data structured in a table of lists form. Each line in the CSV file corresponds to a row in the table. Within a line, commas separate fields, and each field belongs to one table column. CSV files are often used for moving tabular data between two different computer programs (like moving between a database program and a spreadsheet program).
- **Data:** Data is a value or set of values representing a specific concept or concepts. Data become “information” when analyzed and possibly combined with other data in order to extract meaning, and to provide context. The meaning of data can vary depending on its context. Data includes all data. It includes, but is not limited to, 1) geospatial data 2) unstructured data, 3) structured data, etc.
- **/Data page:** A hub for data discovery, which provides a common location, that lists and links to an organization’s datasets. Such a hub is often located at www.example.com/data.
- **Dataset:** A dataset is an organized collection of data. The most basic representation of a dataset is data elements presented in tabular form. Each column represents a particular variable. Each row corresponds to a given value of that column’s variable. A dataset may also present information in a variety of non-tabular formats, such as an extensible mark-up language (XML) file, a geospatial data file, or an image file, etc.
- **/Developer page:** This is a hub for API discovery, which provides a common location where an organization’s APIs and their associated documentation. Such a hub is often located at www.example.com/developer.



- **Database:** A collection of data stored according to a schema and manipulated according to the rules set out in one Data Modeling Facility.
- **Endpoint:** An association between a binding and a network address, specified by a URI, that may be used to communicate with an instance of a service. An end point indicates a specific location for accessing a service using a specific protocol and data format.
- **Error Response Code:** Errors are an inevitable part of API integration, and providing not only a robust set of clear and meaningful API error response codes, but a clear listing of these codes for developers to follow and learn from is essential. API errors are directly related to frustration during developer integration, the more friendly and meaningful they are, the greater the chance a developer will move forward after encountering an error. Put a lot of consideration into your error responses and the documentation that educates developers.
- **GitHub:** GitHub is a social coding platform allowing developers to publicly or privately build code repositories and interact with other developers around these repositories—providing the ability to download or fork a repository, as well as contribute back, resulting in a collaborative environment for software development.
- **Hackathon:** This is an event in which computer programmers and others in the field of software development, like graphic designers, interface designers, project managers and computational philologists, collaborate intensively on software projects. Occasionally, there is a hardware component as well. Hackathons typically last between a day and a week in length. Some Hackathons are intended simply for educational or social purposes, although in many cases the goal is to create usable software. Hackathons tend to have a specific focus, which can include the programming language used, the operating system, an application, an API, the subject, or the demographic group of the programmers. In other cases, there is no restriction on the type of software being created.
- **Information:** Information means any communication or representation of knowledge such as facts, data, or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms.
- **Information Life Cycle:** Information life cycle means the stages through which information passes, typically characterized as creation or collection, processing, dissemination, use, storage,



and disposition.

- **Information System:** Information system means a discrete set of information resources organized for the collection, processing, maintenance, transmission, and dissemination of information, in accordance with defined procedures, whether automated or manual.
- **Information System Life Cycle:** Information system life cycle means the phases through which an information system passes, typically characterized as initiation, development, operation, and termination.
- **JSON:** JSON (JavaScript Object Notation) is a lightweight data-interchange format. It is easy for humans to read and write. It is easy for machines to parse and generate. It is based on a subset of the JavaScript Programming Language, Standard ECMA-262 3rd Edition - December 1999. JSON is a text format that is completely language independent but uses conventions that are familiar to programmers of the C-family of languages, including C, C++, C#, Java, JavaScript, Perl, Python, and many others. These properties make JSON an ideal data-interchange language.
- **JSONP:** JSONP or “JSON with padding” is a JSON extension wherein the name of a callback function is specified as an input argument of the underlying JSON calls itself. JSONP makes use of runtime script tag injection.
- **Machine-Readable File:** This refers to information or data that is in a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost.
- **Metadata:** To facilitate common understanding, a number of characteristics, or attributes, of data are defined. These characteristics of data are known as “metadata”, that is, “data that describes data.” For any particular datum, the metadata may describe how the datum is represented, ranges of acceptable values, its relationship to other data, and how it should be labeled. Metadata also may provide other relevant information, such as the responsible steward, associated laws and regulations, and access management policy. Each of the types of data described above has a corresponding set of metadata. Two of the many metadata standards are the Dublin Core Metadata Initiative (DCMI). The metadata for structured data objects describes the structure, data elements, interrelationships, and other characteristics of information, including its creation, disposition, access and handling controls, formats, content,



and context, as well as related audit trails. Metadata includes data element names (such as Organization Name, Address, etc.), their definition, and their format (numeric, date, text, etc.). In contrast, data is the actual data values such as the “US Patent and Trade Office” or the “Social Security Administration” for the metadata called “Organization Name”. Metadata may include metrics about an organization’s data including its data quality (accuracy, completeness, etc.).

- **OAuth:** This is an open standard for authorization. It allows users to share their private resources stored on one site with another site without having to hand out their credentials, typically username and password.
- **Open Source Software:** Computer software that is available in source code form: the source code and certain other rights normally reserved for copyright holders are provided under an open-source license that permits users to study, change, improve and at times also to distribute the software. Open source software is very often developed in a public, collaborative manner. Open source software is the most prominent example of open source development and often compared to (technically defined) user-generated content or (legally defined) open content movements.
- **Open Standard:** A standard developed or adopted by voluntary consensus standards bodies, both domestic and international. These standards include provisions requiring that owners of relevant intellectual property have agreed to make that intellectual property available on a non-discriminatory, royalty-free or reasonable royalty basis to all interested parties.
- **Parameter:** A special kind of variable, used in a subroutine to refer to one of the pieces of data provided as input to the subroutine. The language defines the semantics for how parameters can be declared and how the arguments get passed to the parameters of subroutines, but the details of how this is represented in any particular computer system depend on the calling conventions of that system.
- **RDF:** Resource Description Framework is a family of specifications for a metadata model. The World Wide Web Consortium (W3C) maintains the RDF family of specifications. The RDF metadata model is based upon the idea of making statements about resources in the form of a subject-predicate-object expression...and is a major component in what is proposed by the W3C’s Semantic Web activity: an evolutionary stage of the World Wide Web in which automated software can store, exchange, and utilize metadata about the vast resources of the



Web, in turn enabling users to deal with those resources with greater efficiency and certainty. RDF's simple data model and ability to model disparate, abstract concepts has also led to its increasing use in knowledge management applications unrelated to Semantic Web activity.

- **REST:** A style of software architecture for distributed systems such as the World Wide Web. REST has emerged as a predominant Web service design model. REST facilitates the transaction between web servers by allowing loose coupling between different services. REST is less strongly typed than its counterpart, SOAP. The REST language is based on the use of nouns and verbs, and has an emphasis on readability. Unlike SOAP, REST does not require XML parsing and does not require a message header to and from a service provider. This ultimately uses less bandwidth.
- **RSS:** A family of web feed formats (often dubbed Really Simple Syndication) used to publish frequently updated works — such as blog entries, news headlines, audio, and video — in a standardized format. An RSS document (which is called a “feed,” “web feed,” or “channel”) includes full or summarized text, plus metadata such as publishing dates and authorship.
- **Schema:** An XML schema defines the structure of an XML document. An XML schema defines things such as which data elements and attributes can appear in a document; how the data elements relate to one another; whether an element is empty or can include text; which types of data are allowed for specific data elements and attributes; and what the default and fixed values are for elements and attributes. A schema is also a description of the data represented within a database. The format of the description varies but includes a table layout for a relational database or an entity-relationship diagram. It is method for specifying constraints on XML documents.
- **SDK:** Software Development Kits (SDK) is the next step in providing code for developers, after basic code samples. SDKs are more complete code libraries that usually include authentication and production ready objects that developers can use after they are more familiar with an API and are ready for integration. Just like with code samples, SDKs should be provided in as many common programming languages as possible. Code samples will help developers understand an API, while SDKs will actually facilitate their integration of an API into their application. When providing SDKs, consider a software licensing that gives your developers as much flexibility as possible in their commercial products.



- **Service-Oriented-Architecture:** SOA expresses a software architectural concept that defines the use of services to support the requirements of software users. In a SOA environment, nodes on a network make resources available to other participants in the network as independent services that the participants access in a standardized way. Most definitions of SOA identify the use of Web services (using SOAP and WSDL) in its implementation. However, one can implement SOA using any service-based technology with loose coupling among interacting software agents.
- **SOAP:** SOAP (Simple Object Access Protocol) is a message-based protocol based on XML for accessing services on the Web. It employs XML syntax to send text commands across the Internet using HTTP. SOAP is similar in purpose to the DCOM and CCORA distributed object systems, but is more lightweight and less programming-intensive. Because of its simple exchange mechanism, SOAP can also be used to implement a messaging system.
- **Swagger:** A specification and complete framework implementation for describing, producing, consuming, and visualizing RESTful web services. The overarching goal of Swagger is to enable client and documentation systems to update at the same pace as the server. The documentation of methods, parameters and models are tightly integrated into the server code, allowing APIs to always stay in sync.
- **Terms of Service:** Terms of Service provide a legal framework for developers to operate within. They set the stage for the business development relationships that will occur within an API ecosystem. Terms of Service should protect the API owner's company, assets and brand, but should also provide assurances for developers who are building businesses on top of an API.
- **TSV:** TSV is a simple text format for a database table. Each record in the table is one line of the text file. Each field value of a record is separated from the next by a tab stop character. It is a form of the more general delimiter-separated values format.
- **Unstructured Data:** Unstructured data is more free form, such as multimedia files, images, sound files, or unstructured text. Unstructured data does not necessarily follow any format or hierarchical sequence, nor does it follow any relational rules. Unstructured data refers to masses of (usually) computerized information that do not have a data structure that is easily readable by a machine. Examples of unstructured data may include audio, video and unstructured text such as the body of an email or word processor document. Data mining techniques are used to find patterns in, or otherwise interpret, this information. Merrill Lynch estimates that more than



85 percent of all business information exists as unstructured data – commonly appearing in e-mails, memos, notes from call centers and support operations, news, user groups, chats, reports, letters, surveys, white papers, marketing material, research, presentations, and Web pages (“The Problem with Unstructured Data.”)

- **Web Service:** A Web service is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format (specifically WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards.
- **WSDL:** An XML-based language (Web Services Description Language) used to describe the services a business offers and to provide a way for individuals and other businesses to access those services electronically.
- **XML:** Extensible Markup Language (XML) is a flexible language for creating common information formats and sharing both the format and content of data over the Internet and elsewhere. XML is a formatting language recommended by the World Wide Web Consortium (W3C).