

A Web Services Taxonomy: not all about the data.

v.1.1, 2009-05-22

Tim McCormick, OCLC Online Computer Library Center, mccormit@oclc.org

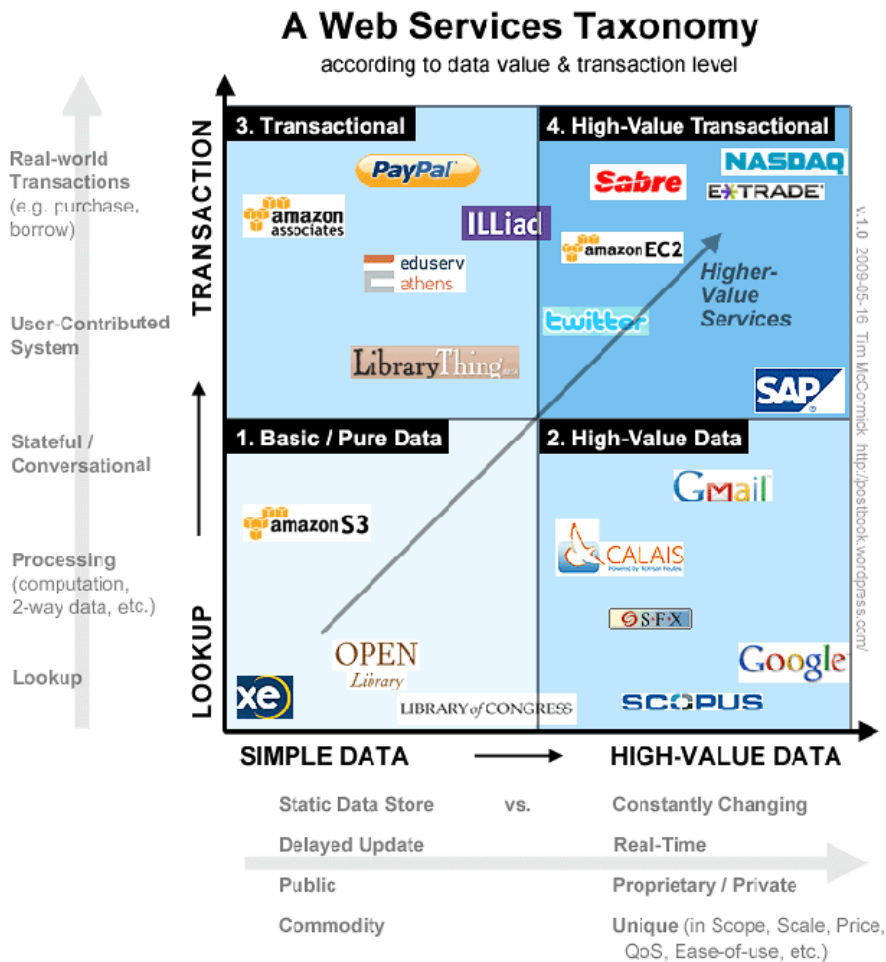


Fig. 1: A Web Services Taxonomy. (see Appendix I for icon key, Appendix II for definition of terms).

Abstract

A Web Service, according to the standard W3C definition, is "a software system designed to support interoperable machine-to-machine interaction over a network." (1)

But, beyond this generic definition, it is useful to consider what are the major types of Web Services in use today, and what properties and business value may be associated with each.

Therefore, I propose a simple taxonomy of Web Services, constructed by characterizing all services according to two factors:

- A) Data quality: from simple/commodity to complex/unique
- B) Transaction level: from basic lookup to real-world transaction.

I give representative examples of Services that exhibit varying degrees of these characteristics. Based on this, I suggest that Services with the most usage, customer value, and/or revenues typically have more complex/unique data, and/or are more transactional.

I. Introduction

A simple taxonomy of Web Services can be made by characterizing all services according to two factors:

- A) Data quality: from simple/commodity to complex/unique
- B) Transaction level: from basic lookup to real-world transaction.

If we represent these two factors by the X and Y axes of a graph, as in Fig. 1, then we can represent any Web Service as a point on the graph, its position determined by our assessment of its relative data quality (X) and transaction level (Y).

So, if we divide this graph into the usual four quadrants, we get four loose categories of services:

1. **Basic / Pure Data Services:** relatively simple, static, and public data; no transactions or statefulness.
2. **High-Value Data:** still no transactions or state, but the data is comparatively complex, proprietary, private, or fast-changing.
3. **Transactional:** service is stateful, or forms a user-generated system, or powers external transactions such as purchases.
4. **High-Value Transactional:** service is transactional, and also provides high-value data.

II. Characterizing Data Quality

Public / commodified data is publicly available by law or de facto, or is easy to get from multiple sources. Often, this coincides with data that is inherently static or is historical (i.e. non-current). For example, a service which returns the current time for a specified time-zone: it uses public, unchanging data, and could be easily implemented by many parties.

Much bibliographic data is comparatively public, because

1. it is generally static;
2. it is considered factual data, and therefore non-copyrightable in the U.S.;
3. many producers of the data, i.e. publishers, have an incentive to freely distribute it;
4. there are many freely-available sources such as the Library of Congress, or Amazon's Product Advertising API (formerly Associates API).

Therefore, to achieve high value, a bibliographic data service must be differentiated by the sum of its quality factors such as scope, currency (updatedness), accuracy, completeness, or ease of use.

Proprietary data is owned or controlled by one party. In some cases, e.g. financial information -- proprietary research, real-time financial-market data, etc. -- access to the data may be strictly controlled and expensive. In other cases, such as Scopus or Google Search, end users may be allowed open access, and there is a supporting business model (pay per view, contextual advertising, etc.).

A special case of Proprietary data service is **Private** services, in which data is maintained for access only by a specific user or group. E.g., online backup and webmail services, and ASP business services such as Salesforce or hosted SAP solutions.

A variation on Private data service is **Contextual** data service, in which the Web Service returns results that are customized to the requestor, based on information in a stored profile or provided at time of request. An example of this is Link Resolver services (aka OpenURL resolvers, e.g. SFX, or WorldCat Link Manager), or proposed Social Search services.

III. Characterizing Transaction Level

The most basic transaction level is a **Lookup** service type in which a query is matched against the key(s) of a database. (as opposed to calculation or processing, in which other and typically more expensive algorithmic steps are required to return a result). These services are usually cases of **One-way Data**, i.e. primary data moves only from service to requestor.

Moving upwards in complexity from Lookup is a group I'll call **Processing**, meaning Services where at a minimum the query may undergo complex transformation or enrichment before

Some examples of this "processing" transaction level are: natural-language processing; enrichment from 3rd-party metadata stores such as CrossRef; query expansion and correction via thesauri, controlled vocabularies, spell-correction, or social data.

In the case of **Two-way Data** the requestor not only receives but supplies primary data, e.g. provides a file to be spell-checked or decoded. For example, Open Calais.

An important special case of two-way data service is a **User-contributed System**. This is an online system in which user interactions and contributions provide some or all of the system's core data or services; e.g., Facebook or Twitter. In some sense, the Wikipedia organization is a case of a User-contributed System, in that the staff and systems it maintains are just a framework for coordinating the contribution and editing activities of its users.

Finally, we call a **Transactional** service on that is capable of initiating or affecting changes in other systems or resources, e.g. a purchase or order or reservation system.

Of course, this is not a rigorous definition, because any functioning system operates by transactions of some sort, and there may not be an clear distinction between the Web Service system and the "other" system that is said to be affected. Nonetheless, we can easily see the commensensical distinction between, say, getting search results vs. booking a plane ticket.

IV. Pure Data services vs. High-value Services

After consider the above cases for contrast, we can further define a **Pure Data service** as: *a web service providing only lookup or search queries, that returns data but is not contextual, stateful, or transactional.*

Hypothesis:

Most highly-valued Web Services are significantly stateful, transactional, and/or non-public. Or, if they are data services, they have best-of-breed scope, quality, or ease-of-use.

Appendix I - Services listed in the taxonomy diagram

1. Data

- **XE** - xe.com, a currency-conversion web service.
- **Open Library** - an open-content library project and archive affiliated with the Internet Archive.
- **Library of Congress** - LC data services such as LC Subject Headings (LCSH) made available as Linked Data.
- **Amazon S3** - Amazon's Simple Storage Solution, a web service providing scaleable online file storage.

2. High-Value Data

- **Gmail** - Google web mail service.
- **Calais** (i.e. Open Calais) - semantic data extraction service from Thomson Reuters.
- **SFX** - OpenURL link-resolver product from Ex Libris.
- **Scopus** - search engine from Elsevier, indexing scientific, technical, & medical literature.
- **Google** - Google.

3. Transactional

- **Paypal** - online payment service, acquired by eBay.
- Amazon Associates (now Amazon Product Advertising API) - Amazon API widely used to get product and bibliographic data.
- **ILLiad** - Interlibrary Loan service from OCLC.
- **Eduserv Athens** - authentication system widely used in the UK to grant access to online academic resources.
- **LibraryThing** - social cataloguing service, which also provides various bibliographic APIs such as ThinglSBN.

4. High-Value Transactional

- **Sabre** - online reservation system for airplane tickets, created by consortium of airlines.
- **Amazon EC2** - Elastic Compute Cloud, a computing-on-demand web service from Amazon.
- **Twitter** - microblogging service and social network.
- **SAP** - producer of enterprise software systems such as ERP (Enterprise Resource Planning) and CRM
- **Etrade** - consumer online service for trading of financial assets.
- **NASDAQ** - online stock quoting and trading system, operated by securities-dealers association of the same name. A purely electronic trading system, in contrast to traditional open-cry exchanges such as NYSE.

Appendix II - Definitions & Notes

This generally follow the W3C's "Web Services Glossary" (2). However, since it is an informal taxonomy I look at common-sensical meanings and implication of the terms, rather than attempting to formally or rigorously define them. Also, I intentionally did not look much for existing Web Service taxonomies, other than Momtahan, et al, 2005 (3), in order to see what my own analysis would come up with.

Conversational: maintaining some limited state between sequential service requests.

Ease of use: the sum of all factors affecting how easily an implementer / user can effectively use a service, as compared to alternate means of achieving the goal. Ease-of-use factors may include sign-up complexity / delay, quality of documentation and support, usability of system design, policy constraints and policy uncertainty, etc.

Lookup: service type in which a query is matched against the key(s) of a database. (as opposed to calculation or processing, in which other and typically more expensive algorithmic steps are required to return a result).

Private: data which a provider maintains for access only by a specific user or group.

One-way Data: usage in which primary data moves only from service to requestor.

Proprietary: data owned or controlled by one party. May be restricted-access or not.

Public data: data that is publicly available by law or de facto, or which is easy to get from multiple sources; it is therefore also commodified.

Pure Data service: web service providing only lookup or search queries, that returns data but is not stateful or transactional.

QoS: quality of service. Technical measures of service quality, typically including service response time, uptime, and throughput. A provider may give customers an assurance of QoS in the form of a **Service Level Agreement**, or **SLA**.

Stateless vs. Stateful:

Transactional: capable of initiating or affecting changes in other systems or resources, e.g. a purchase or order or reservation system.

Two-way Data: usage in which requestor supplies primary data for processing or upload.

User-contributed System: online system, accessed by a Web Service, in which user interactions and contributions make up some or all of the system's value.

(1) "Web Services Architecture" Sec. 1.4. W3C Working Group Note 11 February 2004. <http://www.w3.org/TR/2004/NOTE-ws-arch-20040211/>

(2) "Web Services Glossary." W3C Working Group Note, 11 February 2004. <http://www.w3.org/TR/2004/NOTE-ws-gloss-20040211/>.

(3) "A taxonomy of web services using CSP" by Lee Momtahan, Andrew Martin and A. W. Roscoe. International Workshop on Web Languages and Formal Methods (WLFM 2005). Preprint at http://se10.comlab.ox.ac.uk:8080/LeeMomtahan/docs/momtahan_wlfm05.pdf