The Chaotic Database

Victoria E Holt

Database systems are chaotic by nature. There is no core design, development method, suitable technology or order. They provide no consolidated view of the data and due to the fractal nature of the web can be used seemingly at random.

MySQL AB (M G Mickos – Open Source DB firm) announced at the **2006 Web 2.0** summit that they wanted to create a "Database in the Sky" a distributed repository containing the world's data which is currently held in a structured database. This repository would make structured data available to application developers and internet entrepreneurs where as access to unstructured data is made possible through search engines like Google, see figure 1.

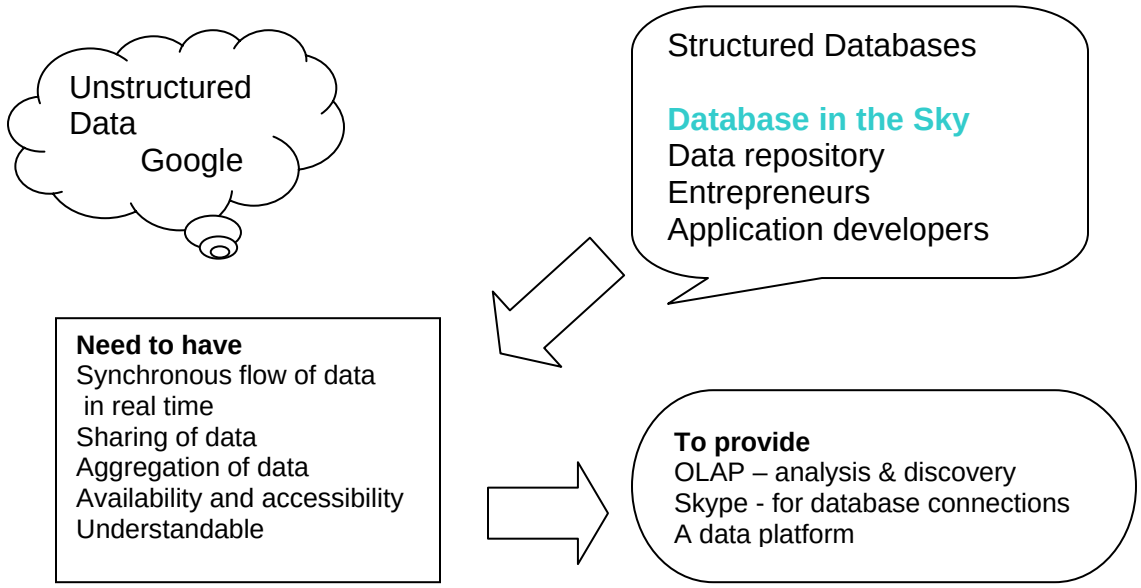

Figure 1 The future vision of the database

The output from this "Database in the Sky" would provide Online Analytical Processing (OLAP), that is to say it provides the ability to read data in databases, create multidimensional data sets and mine the cubes for patterns of data. Skype for database connections follows the Skype technology which allows a peer-to-peer network connectivity which is provided by all the combined users of the software application. This could bring together data like all the weather databases to help analyse the chaotic systems. The database in the sky would require an address system for routing traffic, need to make data definitions understandable and have agreement on what information is shared with permission of the owners. Mickos from MySQL proposed this as an open source project which would need to address some technical challenges.This will need the creation of a

brokerage-like service in the form of trading of who has what data and in what format supported by trustworthy software.

To begin to think of a global linked database, it is necessary to look holistically at the web situation. There are attempts at trying to produce standards but the two worlds of the database and the semantic web, have not quite met. Such items as securing personal data, a data framework and determining what data is to be shared are also addressed by Web 2.0. (T O'Reilly – 2nd generation of Internet Services)

Web 2.0 is emerging as a proposed second generation of internet based services. Web 2.0 includes social networking sites, wikis, communication tools etc that emphasize online collaboration and sharing among users. The name Web 2.0 has been adopted as the popular buzzword for this internet service. The key principles of Web 2.0 applications are:

- the web as a platform
- data as the driving force
- network effects created by an architecture of participation
- innovation in developers e.g. a kind of "open source"
- lightweight business models enabled by content and service syndication
- the end of the software adoption cycle ("the perpetual beta")

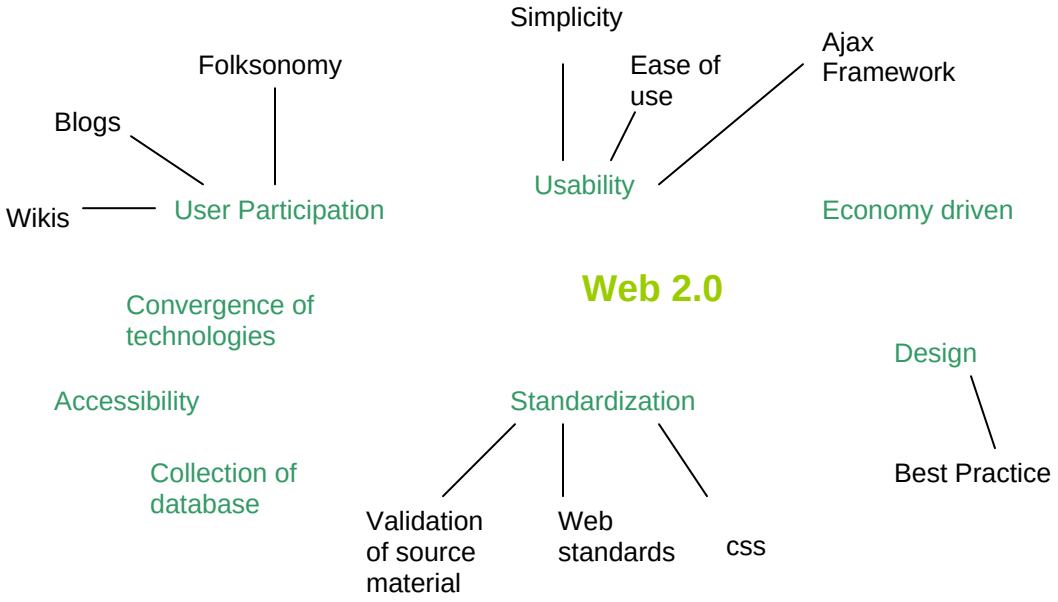

Figure 2 Web 2.0

This suggestion of a Web 2.0 vision, figure 2, is very complex because so many different ways are used to reach the end result. The incentive of sharing should

be enough to encourage people to participate but rarely is this sufficient to motivate people. There were a number of enquiries made about Web2.0 requesting a definition and asking what Web2.0 was about (Wikipedia)

The chaos and complexity of our pluralistic society is being partly addressed by the World Wide Web consortium (W3C – T Berners-Lee) in order to allow databases to become a part of this global vision as the semantic web. This web of data as defined by the W3C, provides a common framework that allows data to be shared and used across applications, enterprises and community boundaries. The database in the sky debate is a similar iteration of Tim Berners-Lee's vision in 1998. For this great database to happen more people have to adopt a common model for social and technical requirements. This problem is both of a hard and soft system making indicating that the resolution and option process are hard to achieve.

A list of the attributes are as follows:

1. a social model for data creation exploiting the network effect
2. intentional data exposure for recombination and consumption
3. agreements on common data formats
4. addressability of data; Universal Resource Idententifiers (URI's) identify resources and are central to the semantics web global naming convention. It is necessary to have a formal definition of terms so that everyone identifies objects with the same terminology throughout the world.
5. pervasive networking e.g. a directory of data collections and services using web services to route straight into that content.
6. cross reference concepts between systems
7. licensing of data
8. availability
9. accessibility
10. proof of where the data originate
11. encryption of sensitive data

The architecture of the semantic web is displayed in figure 3.

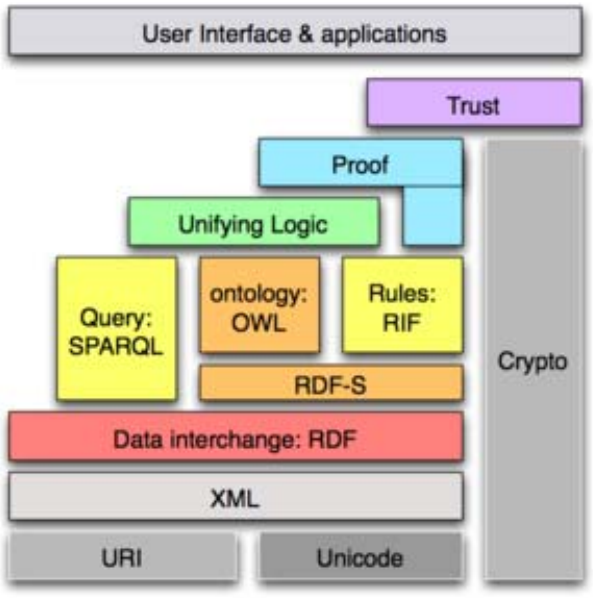

Figure 3 Semantic Web Architecture (ver. 2006)

Adopted from Wikipedia,
(Web layer cake)
the free encyclopedia

The W3C have a framework to address some of the issues of the web of data which requires a new global query language like XML. The Resource Description Framework (RDF) is used to represent information and to exchange knowledge in the web. XML is the language for syntax whereas RDF is for semantics. RDF is basically a clear set of rules for providing sample descriptive information. The OWL Web ontology language is used to publish and share sets of terms called ontologies supporting advanced web search, software agents and knowledge management. Subject or domain specific vocabularies ontology define the terms used to describe and represent an area of knowledge. This helps organise knowledge.

The complexity of deep ontologies has led some people to take an altogether different approach. Folksonomies are a development generating a considerable interest at the moment. They represent a structure that emerges organically when individuals manage their own information requirements. Folksonomies are created when a large number of people are interested in particular information. The groups are encouraged to describe it in full and to then create a set of standard tags. This large communities practice is evolving slowly.

The semantic web is about creating a global medium of exchange of information via the use of standards, markup and related processing tools. The vision is of web pages that are understandable by computers so cross website searching and performing actions can be standardized. Currently all this data exists but the computers don't know what it is and how it should be interrelated. It is hoped it will enrich our lives by the data becoming accessible so that we can write programs that help to understand this globally distributed data. The computer needs some human direction. This simple idea remains largely unrealized to date and this leading edge technology is yet to be addressed.

Communities will find a way forward by either adopting or not adopting methodologies. Work in practice and conceptualization of terminologies in socially subtle ways will influence development and maintenance will change. The web of databases exposes another type of database called the data space which consists of raw data, information (data in context), web services and increasing knowledge. A dataspace has all the functionality of a database with additional benefits of being a data model and query language agnostic.

The web and database's fractal nature become more and more apparent the deeper it is investigated.

"We expect the developments, methodologies, challenges and techniques we've discussed here to not only give rise to a new semantic web but also contribute to a new web science – a science that seeks to develop, deploy, and understand distributed information systems, systems of humans and machines, operating on a global scale. Artificial Intelligence (AI) will be one of the contributing disciplines"
In order to make learning machines out of the current 'pattern-match and calculation' machines
The semantic web revisited May/June 2006 www.computer.org/intelligent

So where does that leave us in the world of the systems? Well this demonstrates that the problem will not be solved by technology alone but by the analysis of these complex systems.

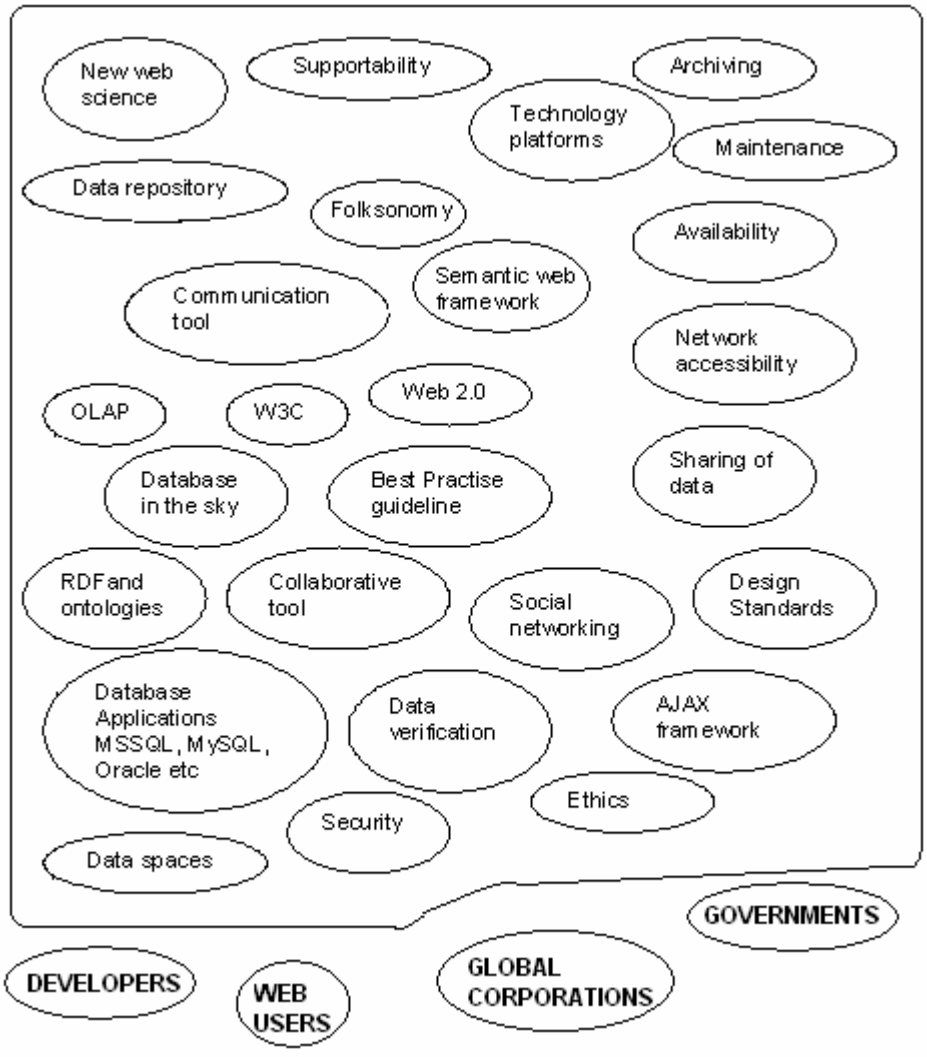

Figure 4 Systems map of the Chaotic Database

The systems map shows all the main elements involved within this complex system. People have different perspectives and are eclectic in nature which

results in the situations being influenced by each other thus adding layers of complexity to this already complex situation. Every participant has their own agenda and specific design projects which are more often than not carried out in isolation from the bigger picture. This results in a cause and effect scenario on the database. The web is a global tool affected by its environment where anything is possible but our epistemological awareness required to achieve emergence of the database system requires systems thinking in practice to create a holisitic vision for the future. I will cover this in more detail in another paper.

References.

MySQL wants to build 'database in the sky' By Juan Carlos Perez, IDG News Service November 09, 2006 http://www.infoworld.com/article/06/11/09/HNdatabaseinsky_1.html

MySQL's Mickos shares 'database in the sky' vision November 9, 2006 12:32 PM
http://news.com.com/2061-10795_3-6134142.html

The great [structured] database in the sky November 9th, 2006
http://blogs.zdnet.com/BTL/?p=3918

The Great Database In The Sky
Posted by davidvc on 13 Nov 2006 at 05:34 PM David Van Couvering Blog
http://weblogs.java.net/blog/davidvc/archive/2006/11/the_great_datab.html

Internet Alchemy - The Great Database in the Sky
http://iandavis.com/blog/2006/11/the-great-database-in-the-sky

Web Of Data Posted by Ian Davis at May 8, 2006 10:21 AM
http://blogs.talis.com/nodalities/2006/05/web_of_data.php

Semantic Web http://www.w3.org/2001/sw/

The Semantic Web Revisited, by Nigel Shadbolt, Tim Berners-Lee and Wendy Hall, IEEE Intelligent Systems 21(3) pp. 96-101, May/June 2006

The Semantic Web: an Interview with Tim Berners-Lee, by Andrew Updegrove at ConsortiumInfo.org, June 2005

'The Semantic Web lifts off' by Tim Berners-Lee and Eric Miller, W3C. ERCIM News No. 51, October 2002.

The Semantic Web, Scientific American, May 2001, Tim Berners-Lee, James Hendler and Ora Lassila

World Wide Web Consortium Issues RDF and OWL Recommendations
http://www.w3.org/2004/01/sws-pressrelease

Web 2.0 From Wikipedia, the free encyclopedia
http://en.wikipedia.org/wiki/Web_2.0

Semantic Web From Wikipedia, the free encyclopedia

http://en.wikipedia.org/wiki/Semantic_Web

Skype http://en.wikipedia.org/wiki/Skype
Essential SQL Server 2000 Administration Handbook, Buck Woody