

June 2010

# "The Semantic Web and Online General Reference -- Are We There Yet, or Any Time Soon?"

John G. Dove

*Credo Reference*, johngdove@gmail.com

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

---

### Recommended Citation

Dove, John G. (2010) ""The Semantic Web and Online General Reference -- Are We There Yet, or Any Time Soon?,"" *Against the Grain*: Vol. 22: Iss. 3, Article 15.

DOI: <https://doi.org/10.7771/2380-176X.5554>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

# “The Semantic Web and Online General Reference — Are We There Yet, or Any Time Soon?”

by **John G. Dove** (President, Credo Reference) <dove@credoreference.com>

**Column Editor’s Note:** *In this paper I propose an approach to applying the concepts of the Semantic Web to the area of online general reference for libraries. This is motivated by a belief that if content owners can prepare their content in an appropriate fashion for participation in online environments, then user interface developers can provide users with compelling capabilities that go across a very broad spectrum of library-acquired content. This will make that content far more valuable to users and will even mean that user interface developers working for one publisher would have ways of interacting with content from many other publishers and vice versa. This would open up a significant set of opportunities for innovation for the benefit of users, libraries, and the publishers who provide them content.*

(Based on a talk, “The Problem of the Common User Interface,” **Charleston Conference**, November, 2008, presented by **John Dove** and **Robert Scott**.) — **JD**

**“In an information-rich world, the real design problem to be solved is not so much how to collect and distribute more information but rather how to increase the rate at which persons can find and attend to information that is truly of value to them.”**

— **Peter Pirolli**, *Information Foraging Theory*, Oxford University Press, 2007

**Tim Berners-Lee**, **James Handler** and **Ora Lassila** laid out a clear vision in the May 17, 2001 issue of *Scientific American*, of how the Semantic Web could truly change our lives. Just over seven years before then, **Tim Berners-Lee**, the inventor of the World Wide Web, provided the world with some of the first WWW information servers and started the remarkable explosion of practical uses of the Internet.

It is now seven years since this Semantic Web paper was published and it’s fair to ask, how much has the vision affected the world of Online General Reference in libraries which is the world that we are both involved with, myself as president of **Credo Reference**, and **Bob** as the Head of Electronic Text Services for **Columbia University Libraries**?

The 2001 Semantic Web paper painted a very concrete vision of one way in which daily life’s complexities could be solved more easily in a world where information was identifiable to computers by the actual meaning of the information. Here’s the picture the authors painted:

Lucy is with her mother at the doctor’s office having just received word that her Mom has a medical condition that will require appointments with a specialist and physical therapist. A phone call to her husband, Pete, and a few quick interactions on both their parts with computer systems that understand and can take advantage of the Semantic Web, and after many details are sorted out by cooperating computer systems with information from Lucy, Pete, specialists, physical therapists, including calendars, appointments, insurance, as well as evaluations of the quality, location, and availability of various services, Lucy and Pete enjoy the fact that they have a completely worked out schedule and plan for Lucy’s Mom by the time dinner is ready that evening — and with only minimal work on Lucy or Pete’s part.

While many advances in technology, information standards, and ability of applications to take advantage of the semantics of information have now accrued, this vision of Lucy and Pete enjoying the fruits of the Semantic Web is still a good way off.

Still, this vision is very compelling even if it seems as far off today as it was in 2001. The set of cooperating organizations, processes, and practices that would need to successfully adopt semantic web technologies in order for the Lucy/Pete vision to become a reality is actually quite complicated. So perhaps that is part of the reason that this vision is still a good way off. But even in some relatively simple worlds with a small number of organization types, use of the Semantic Web is still only a dream. The world we are interested in is online reference content for libraries. In this world there are a couple hundred publishers, and a few dozen user interface and technology providers, and the libraries which they share as customers.

Even in this simpler and technology-aware domain of online reference, little has been done to implement the Semantic Web.

So I’d like to present a much more modest vision — limited just to how the Semantic Web might revolutionize how reference resources could be used collaboratively in online reference services for libraries of all types, all around the world. I propose this vision in hopes that because it is far simpler than the Lucy/Pete vision, there is perhaps a better chance that it can come to fruition in reasonable time.

Reference resources that make up much of the content of reference rooms in physical libraries do not represent the whole library; rather they are a distillation of material (sometime even referred to as “tertiary” resources) where “primary” is original source

material, “secondary” is material that represents experts’ analysis of the primary source material, and “tertiary” is coverage of the most important facts, definitions, and explanations that will be most useful for patrons of that particular library, letting them get a quick fact or the grounding in the overall landscape of a field they are studying.

More and more of such resources are now online, and it’s pretty easy to see a world in the next few years in which all such resources are online. So here is my vision:

Imagine a certain university-level student, **Judy**. She’s a second-year undergraduate student at **Oberlin College** (where it turns out **Bob** and **I** were both students back in the 1960s). **Judy** is passionate about history, but her current course in women’s studies has introduced her to the field of sociology. So she’s not yet decided on whether her major will be in the humanities or in the social sciences. Of course she read *David Copperfield* and *Great Expectations* in high school, but she never really understood the whole context of the world **Charles Dickens** lived in and wrote about. She’s thinking of doing her next paper for her women’s studies course on **Charles Dickens**. So **Judy** is browsing through her college library’s online resources (which she, remarkably, found via some links from Wikipedia) and ends up on some browse pages on the **Oberlin College English Literature subject portal**. From a biographical article about **Charles Dickens**, **Judy** finds a chronology of his life which makes note of the fact that in the 1830s he lived on Doughty Street in London. Clicking on that link she sees the following **Google Map** locating **Dickens’** home and the whole map of streets anywhere within one mile of this address. *See MAP #1, page 38.*

I’m assuming for the purposes of this vision that the online reference resources available to

this browse page about **Charles Dickens** are some of the standard kinds of resources that exist today in subject encyclopedias from good quality publishers as one would expect to find in the **Oberlin College Library’s** online resources: **Oxford University Press**, **Columbia University Press**, **Cambridge University Press**, **Routledge**, **Gale**, **Macmillan**, **Sage Reference**, **Elsevier**, etc.

If this set of online reference content could

*continued on page 38*



**The Semantic Web ...**  
from page 36

be prepared to capture and share some simple semantics then the following user experience would be more easily achievable. The simple semantics I suggest would be ways of assigning to encyclopedia entries indications of what famous people, places, events, works, or institutions are associated with such entries.

Judy's **Google** map begins to be populated with interesting pieces of content related to **Charles Dickens** and the world he lived in during the early 1800s on Doughty Street.

Famous architects, philanthropists, and key historical events show up on the **Google** map, but none of these catch Judy's interest. *See MAP 2, this page.*

But then she notices the Clerkenwell riots just four blocks from **Dickens** home. *See MAP 3, this page.*

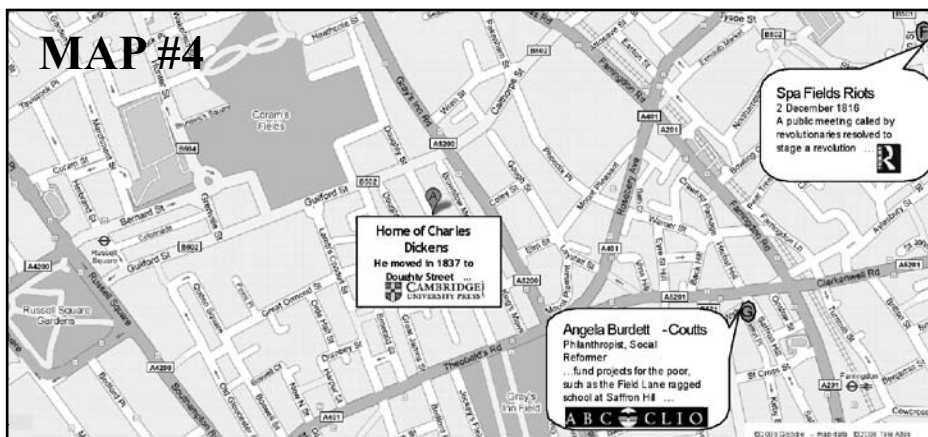
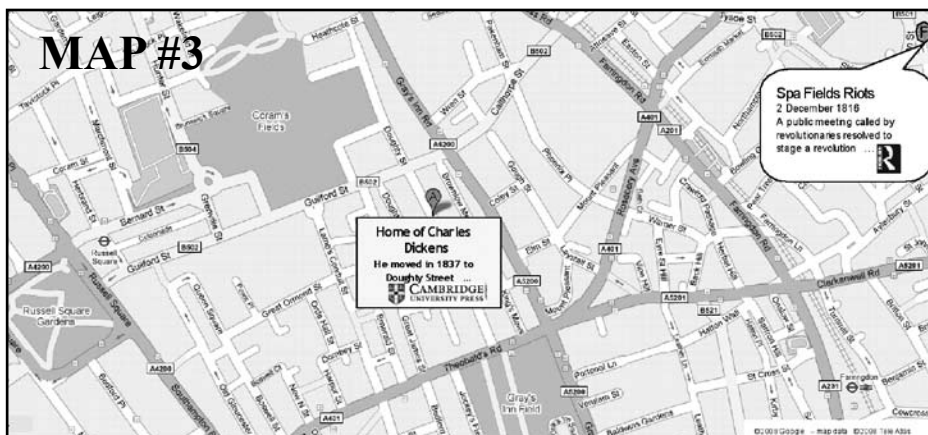
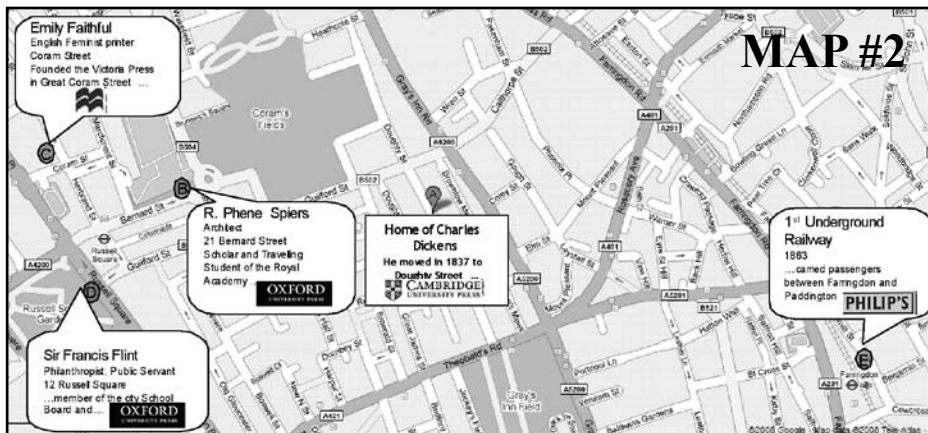
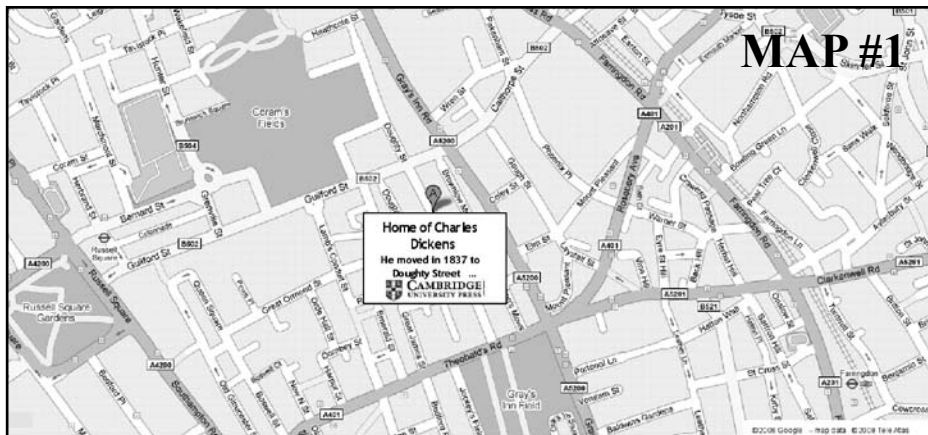
Here, a mass meeting called to press for universal suffrage was broken up by mounted police troops. She is wondering whether or not **Dickens** would have written anything about these Clerkenwell riots when she notices that the social reformer, **Angela Burdett-Coutts** lived nearby and that she is described in an entry from **ABC-CLIO's Encyclopedia of Women Social Reformers** as a friend of **Charles Dickens**. *See MAP 4, this page.*

She thinks this might be her topic, but then she notices the Foundling Hospital just two blocks north of **Dickens**' house. The Foundling Hospital ended up on this map because **Oberlin's** online reference library included *Chambers Biographical Dictionary* which had a short biography of **Thomas Coram**, the hospital's founder. *See MAP 5, page 40.*

This really catches her attention. When mousing over that point on the map, she discovers that this hospital cared for thousands of abandoned babies. This becomes her chosen topic. After clicking through to more information on **Thomas Coram**, she learns that he was a sea captain from Taunton, Massachusetts, just thirty miles south of her home in Boston. The article is linked into **Oberlin College's** other fine set of resources for her further research.

Had Judy continued her browsing to the west she might have noticed an entry about a doctor, **Peter Roget**, who lived a couple blocks west of **Charles Dickens**. Right-clicking on this point of the map now makes the context map she's looking at center itself around **Dr. Peter Roget**. It would not be a stretch to imagine that **Charles Dickens** knew **Peter Roget** because they both frequented the Reading Room at the **British Museum**. And this doctor, late in life, decided others might benefit from a synonym list he constructed over the years for his own use — what became *Roget's Thesaurus*. *See MAP 6, page 40.*

*continued on page 40*



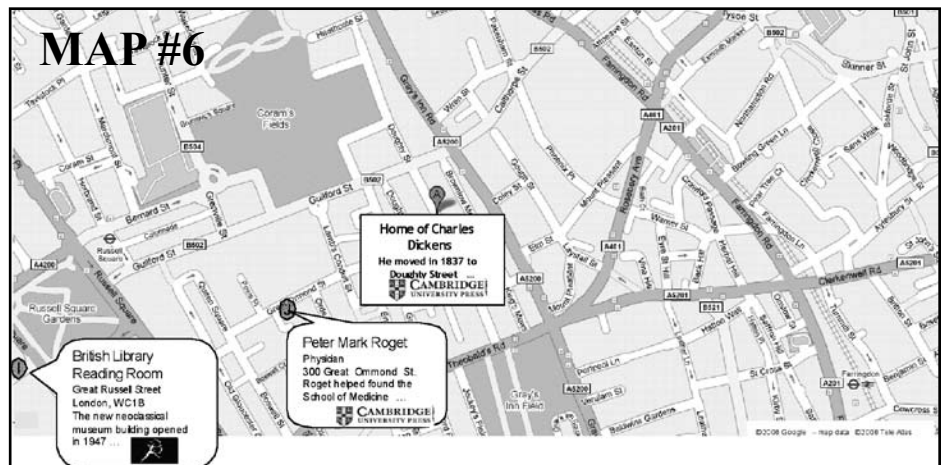
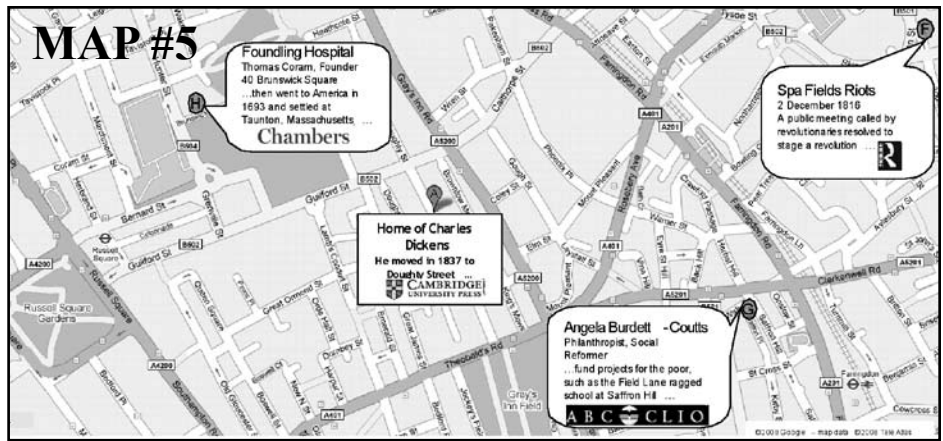
**The Semantic Web ...**  
from page 38

It's worth thinking about what could make this vision work. One is that general reference content is NOT the whole library. One could easily imagine that a good college library might have online databases that include the entire census of that part of London in the 1830s, or descriptions of every single address in that neighborhood. Putting that whole census or listing of addresses onto the **Dickens Google Map** suddenly overwhelms the context and the browsing becomes untenable. This vision of a browsable map takes advantage of the fact that subject encyclopedias (the core content type in general reference) has definitions and explanations of the *important* people, places, events, works, and institutions related to a particular field. And any particular library would appropriately assemble in their online general reference system the set of subject encyclopedias that are relevant to their patrons and users. So the **Dickens Google** map might look very different at the online reference system at a fashion design institute in New York City, or a middle school in Scotland, or a research institute in Triangle Park.

There are some wonderful examples online of what one can do with maps and linking them to information about places. Of particular recent note is the fabulous visualization of the Roman Forum created by **UCLA's Cultural Virtual Reality Laboratory**. This shows that properly prepared information can provide an astounding learning experience. Those premium examples of visualization of information are what **Ronald Milne** of the **British Library** calls "cabinets of curiosity," wonderfully designed information experiences but cut off from each other and from the main stream of information exploration.

What I am painting in this vision is much more modest in functionality but generic across all general reference information for libraries of various types. If the right standards are adopted by leading reference publishers, maps such as this **Charles Dickens** map become universal across a myriad of topics, people, places, events, etc. This does not require any new content development — simply applying Semantic Web technologies to existing online representations of reference works and letting librarians pick and choose which ones participate in their users' experience.

In October 2008, *Scientific American* published another major article on the Semantic Web — "Web Science Emerges" by **Nigel Shadbolt** and **Tim Berners-Lee**. Among other things it describes how **DBpedia**, a project initiated by **Chris Bizer** et al at the **Free University of Berlin**, has applied RDF (Resource Descriptive Framework) technologies of the Semantic



Web to harvest useful semantic information from infoboxes in **Wikipedia**.

Looking to see if this would be a part of what might bring my vision to the light of day, I notice that **DBpedia** has now created over one hundred and ten million RDF triples from infoboxes on **Wikipedia** representing thousands of classes of information. Looking over **DBpedia's** list of several thousand semantic classes, it strikes me that this is both too little and too much for achieving the vision I have for Judy's discovery of **Dickens'** context in the 1830s. Among the classes are "names of female hockey teams in the Southern United States" and "persons who died in 101 B.C." The semantic categories are so numerous and not clearly interrelated, so I don't see how interface developers would use them to easily create things like my example above.

If achieving the vision for Judy means that **Oxford University Press**, **Macmillan**, etc., or someone who aggregates content from them is going to have to indicate the semantics of information in their online encyclopedias, then I think it is imperative that the list of semantic categories be carefully selected and well-thought through. In that way one can imagine the hundred or so top quality publishers of subject encyclopedias are more likely to end up doing the work necessary to indicate the appropriate semantics of their content.

Of course some content owners will want to provide additional detail in identifying the semantics of their content. Biographical sources may want to distinguish among important persons who are artists, scientists, or politicians, etc. Some sources will want to identify publishers or counties or universities, all examples of institutions. But as long as these are defined in a standard manner so they are sub-classes of Persons, Places, Events, Works, or Institutions, then interface developers can work with them across all fields of study.

I suggest that online reference publishers could start with classes of reference content such as Persons, Places, Events, Works, and Institutions, ... and have attributes such as "defined," "explained," "image-of," "produced-by," and "mentioned" .... And on that foundation one can imagine not only the vision I've painted for Judy seeing a **Google** map populated with interesting and potentially insightful bits of information about **Dickens'** world of 1837 in London, but a good number of other fabulous ways to browse the information her institution's librarian assembles into the online reference experience for that library.

I believe that if a simple set of metadata classes can be endorsed by the online reference publishing world, it will enable a whole new class of user experiences that leverage the vast quantities of reference content which is already online. 🐼