

Webinar Overview

- Intro: Schema.org, Linked Data & RDF
- A quick look at schema.org
- A longer look at Linked Data and RDF
- Advantages and Challenges of RDF / Linked Data
- Schema.org in detail
- Combining Linked Data and schema.org
- The Future, and what it may hold

Goals

- Show common origins of schema.org and Linked Data
- Illustrate the different design choices available
 - for identifying real-world entities
 - syntaxes for describing things in structured data
 - for integrating multiple vocabularies
 - for incentives to publish (accurate, detailed) data
 - usability, expressivity and pluralism
- ...helping guide adoption of schema.org + Linked Data

Schema.org

A quick look.



Insights from Googlers into our products, technology and the Google culture



Search

Introducing schema.org: Search engines come together for a richer web

June 2, 2011 at 7:00 PM



(Cross-posted on the Inside Search Blog)

Today we're announcing schema.org, a new initiative from Google, Bing and Yahoo! to create and support a common vocabulary for structured data markup on web pages. With schema.org, site owners and developers can learn about structured data and improve how their sites appear in major search engines. The site aims to be a one stop resource for webmasters looking to add markup to their pages.

Search engines have been working independently to support structured markup for a few years now. We introduced rich snippets to Google search in 2009 to help people find better summaries of reviews and people, and since that time we've expanded to new kinds of rich snippets, including recipes and events. We've been thrilled to see content creators across the web—from stubhub.com to allrecipes.com—add markup to their pages, and today we're able to show rich snippets in search results more than 10 times as often as when we started two years ago.

Connect with us Subscribe to this blog:



FeedBurner



RSS Feed

Browse all of Google's blogs for specific interests & topics:



Blog Directory

Follow Google alerts and news on Twitter:



@Google

We'd love to hear your feedback — discuss Onnelele eneducte with

Blogs Toolbox Media Home Forums Events

Search within the com

One person

first of your

Site Blogs » Search Blog » Introducing Schema.org: Bing, Google and Yahoo Unite to Build the Web of Objects

Search Blog

Introducing Schema.org: Bing, Google and Yahoo Unite to Build the Web of Objects

The Bing Team 6/2/2011 10:01 AM Comments (5)



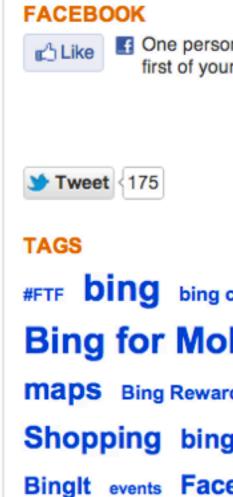
We've been talking for a while about the need to rethink the search experience to better reflect both the changing web and advancing user habits.

One of the biggest challenges and opportunities we see is to literally create a high-definition proxy of the physical world inside of Bing. In other words, we want to be able to model the world in which we all live to the level that search can actually help you make decisions and get things done in real life by understanding all the options the world presents.

We've made great progress on the technical front to begin to model the real world from the messy bits of data scattered across the web. Things like movies have benefitted from this work. We're now able to understand "Casablanca" is a movie and literally mine the web to re-assemble information about that movie from millions of sites.

But we think we can do better. We want to enable publishers to give us hints about what things they are describing on their sites. Rather than rely solely on machine learning and other Al techniques, we asked "what if we could enable publishers to have a single schema they could use to describe their sites that all search engines could understand?"

Well today, we're pleased to announce Bing is joining forces with Google and Yahoo! to deliver schema.org, a new initiative, to create and support a common set of schemas for structured data markup on web pages. With schema.org, site owners and developers can learn about structured data and improve how their sites appear in



instant answers K

live search

home

subscribe | help



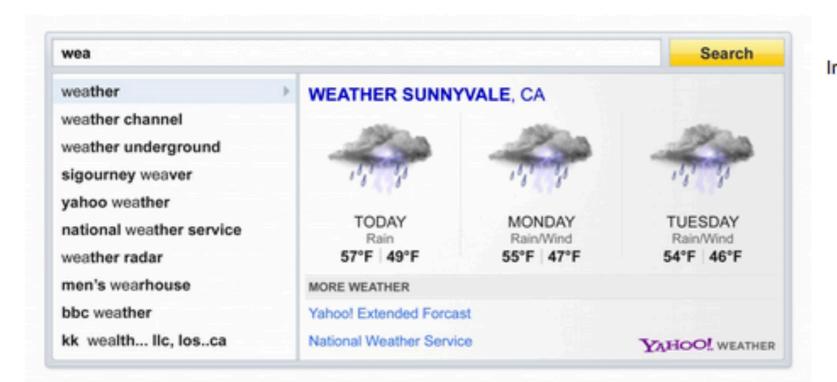
back to yahoo! search

Introducing schema.org: A Collaboration on Structured Data

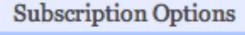
Posted June 2nd, 2011 at 11:50 am by Yahoo! Search

Categories: Search

Today we're announcing schema.org, a new initiative from Yahoo!, Bing, and Google, to create and support a common set of schemas for structured data markup on web pages. With schema.org, webmasters and developers can learn about structured data and improve how their sites appear in search results on Bing, Google, and Yahoo!. Information and tips are available on schema.org, a onestop resource for webmasters looking to add markup to make their pages better understood by search engines.



2008, Yahoo! showed its support of structured data through the launch of the SearchMonkey program. We won plaudits and awards for our work but most importantly Yahoo! helped boost an increase in markup on the web which in turn helped improve the search experience by resolving questions quickly through enhanced result displays. SearchMonkey also helped Yahoo! kick start various initiatives to









RSS

Facebook Fans



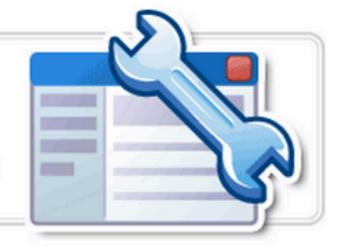
Latest Posts

For example?



Google Webmaster Central Blog

Official news on crawling and indexing sites for the Google index



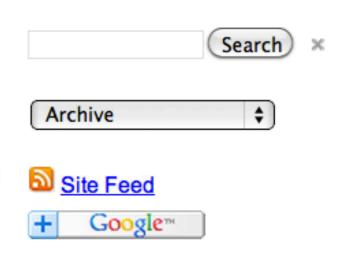
Using schema.org markup for videos

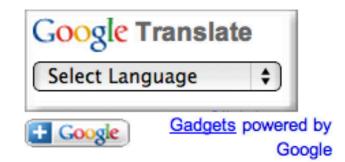
Tuesday, February 21, 2012 at 2:00 PM Webmaster level: All

Videos are one of the most common types of results on Google and we want to make sure that your videos get indexed. Today, we're also launching video support for schema.org. Schema.org is a joint effort between Google, Microsoft, Yahoo! and Yandex and is now the recommended way to describe videos on the web. The markup is very simple and can be easily added to most websites.

Adding schema.org video markup is just like adding any other schema.org data. Simply define an itemscope, an itemtype="http://schema.org/VideoObject", and make sure to set the name, description, and thumbnailUrl properties. You'll also need either the embedURL — the location of the video player — or the contentURL — the location of the video file. A typical video player with markup might look like this:

```
<div itemscope itemtype="http://schema.org/VideoObject">
  <h2>Video: <span itemprop="name">Title</span></h2>
  <meta itemprop="duration" content="T1M33S" />
  <meta itemprop="thumbnailUrl" content="thumbnail.jpg" />
  <meta itemprop="embedURL"
    content="http://www.example.com/videoplayer.swf?video=123" />
  <object ...>
   <embed type="application/x-shockwave-flash" ...>
  </object>
  <span itemprop="description">Video description</span>
</div>
```





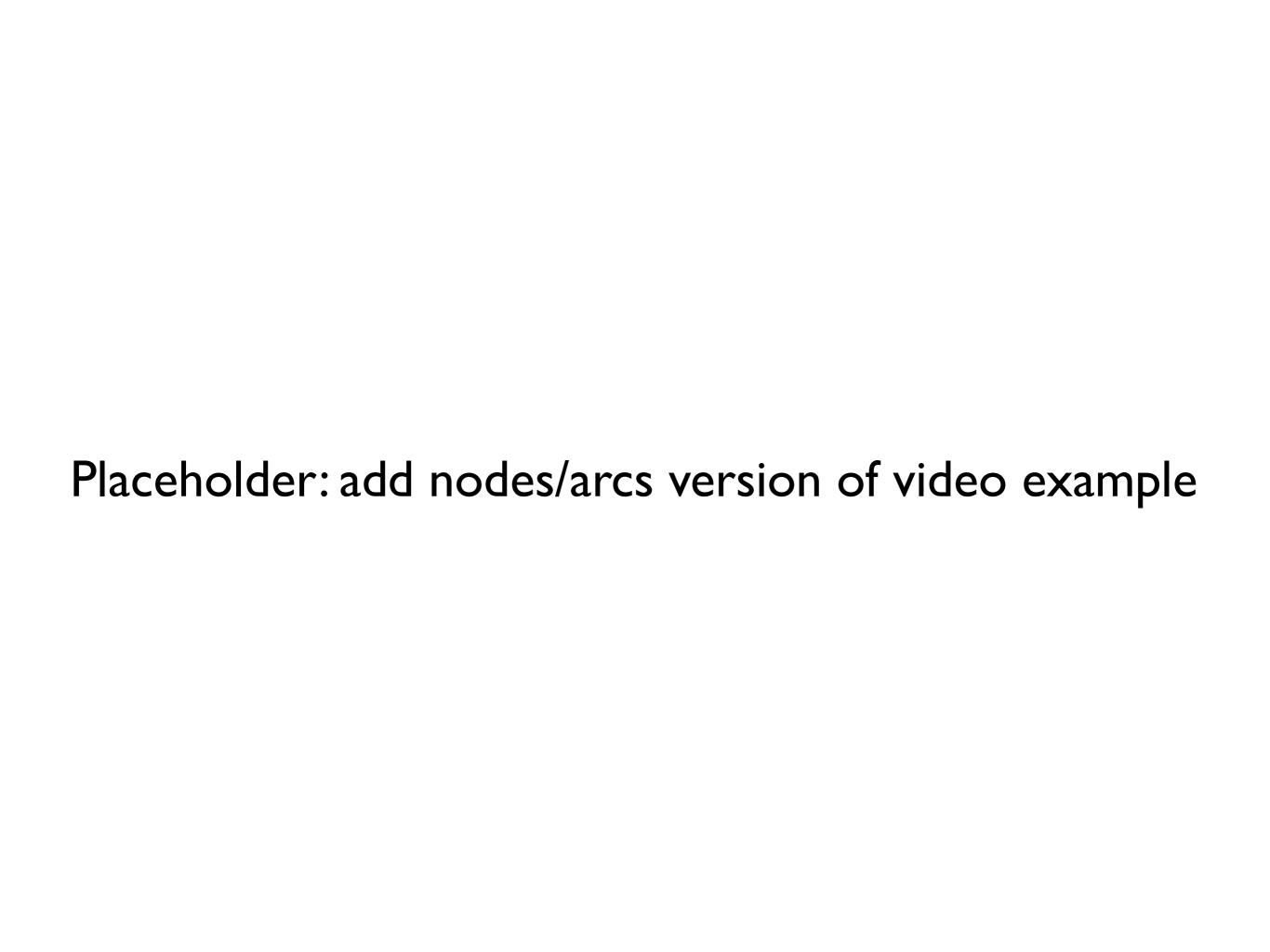
Useful links

Google Webmaster Central Webmaster Help Center Google Webmaster Tools

Webmaster Central on YouTube

```
<div itemscope itemtype="http://schema.org/VideoObject">
    <h2>Video: <span itemprop="name">My Title</span></h2>
    <meta itemprop="duration" content="T1M33S" />
    <meta itemprop="thumbnailUrl" content="thumbnail.jpg" />
    <meta itemprop="embedUrl"
        content="http://example.com/videoplayer.swf?video=123" />
        <object ...>
            <embed type="application/x-shockwave-flash" ...>
            </object>
            <span itemprop="description">Video description</span>
</div>
```

```
Type: http://schema.org/VideoObject
name = My Title
duration = T1M33S
thumbnailurl = thumbnail.jpg
embedurl = http://www.example.com/videoplayer.swf?
video=123
description = Video description
```



How does it work?

- Normal HTML page content, with additional markup.
- Extra attributes are from W3C HTML5 Microdata: itemscope, itemtype, itemprop, content.
- The HTML5 Microdata syntax was based originally on W3C RDFa, a similar design.
- Both address the principle known on microformats.org as "Don't Repeat Yourself" (DRY), http://microformats.org/wiki/dry.
- Re-use of page markup to also serve as structured data.

Linked Data

Where did it come from...

...and where is it going?

Linked Data: Prehistory

- Before going deeper, let's go back to the origins of Linked Data publication strategies.
- 'Linked Data' as a slogan and community began with a Tim Berners-Lee note (July 2006) which inspired a Linking Open Data community group and who in turn inspired many others to publish linked, structured data.
- What was new about the 'Linked Data' recipe? How did it improve upon previous RDF practice 1997-2005? How does it compare to the structured data style of schema.org? How did we move from having only a handful of linked RDF files in 2002, through millions to today's talk of billions?

Linked Data

- I.Use URIs as **names** for things.
- 2. Use **HTTP** URIs so that people can look up those names.
- 3. When someone looks up a URI, **provide** useful information, using the standards (RDF, SPARQL, ...).
- 4. Include **links** to other URIs so they can discover more things.

Tim Berners-Lee (2006)

http://www.w3.org/DesignIssues/LinkedData

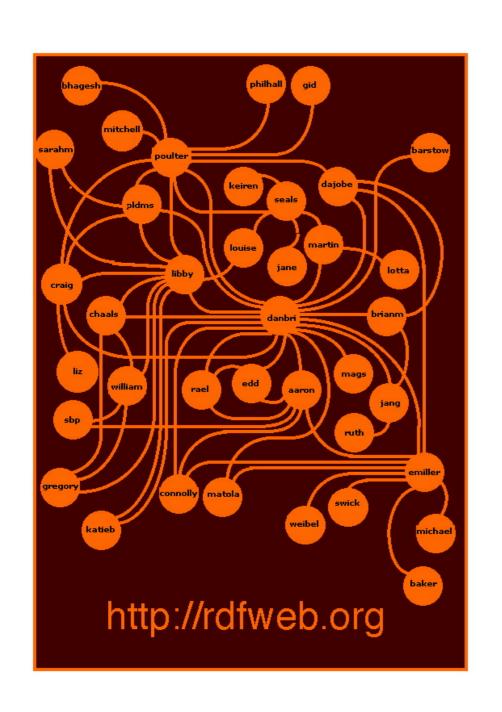
From FOAF to Linked Data

- TimBL's 'Linked Data' note critiqued FOAF's style of linking, since FOAF 2000-2005 did not encourage use of 'http:' identifiers for people and real-world things.
- Instead, FOAF files used a "reference by description" style, describing people in terms of their properties, with "seeAlso" links into a Web of linked RDF documents.
- "This [Friend of a Friend] linking system was very successful, forming a growing social network, and dominating, in 2006, the linked data available on the web." (TimBL 2006)
- "However, the system has the snag that it does not give URIs to people, and so basic links to them cannot be made."
- Post-2007 (with supporting changes in the standards) the RDF community adopted this new style, and started publishing RDF with "http:" URIs for real-world things.

Linked RDF: Growth

- In 2002, an early FOAF crawl found RDF about 24 people.
- By 2006 (see Jennifer Golbeck 'Web-based Social Networks A Survey and Future Directions') we saw:
 - LiveJournal.com: 7,500,000 FOAF profiles
 - Ecademy.com: 72,000 FOAF profiles
 - Trust project: 1,700 FOAF profiles
 - ...others, totalling 6 million+ site-based profiles
 - plus smaller number of "in the wild" FOAF files

2002: all the Linked RDF we crawled could fit on a t-shirt:



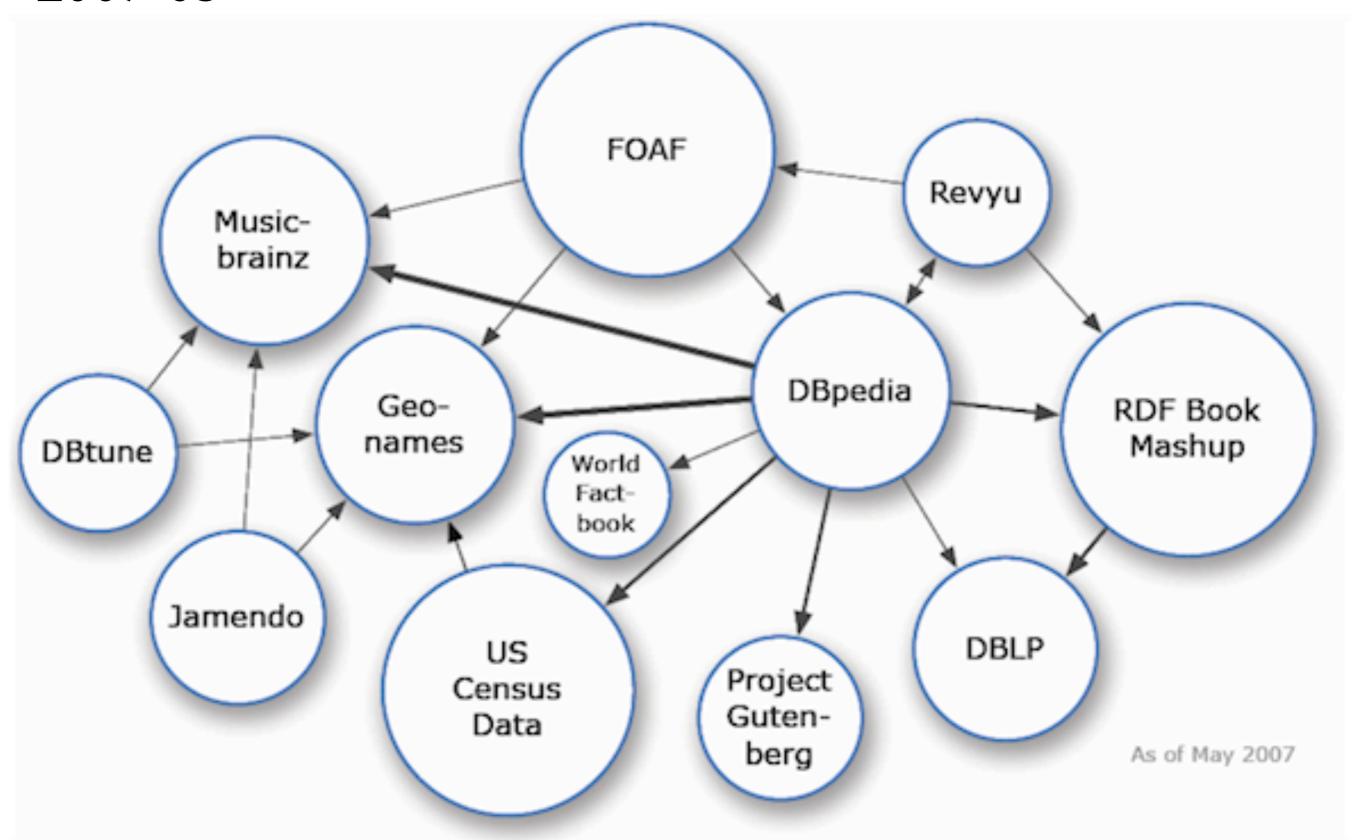
...by 2006 finding millions of files.

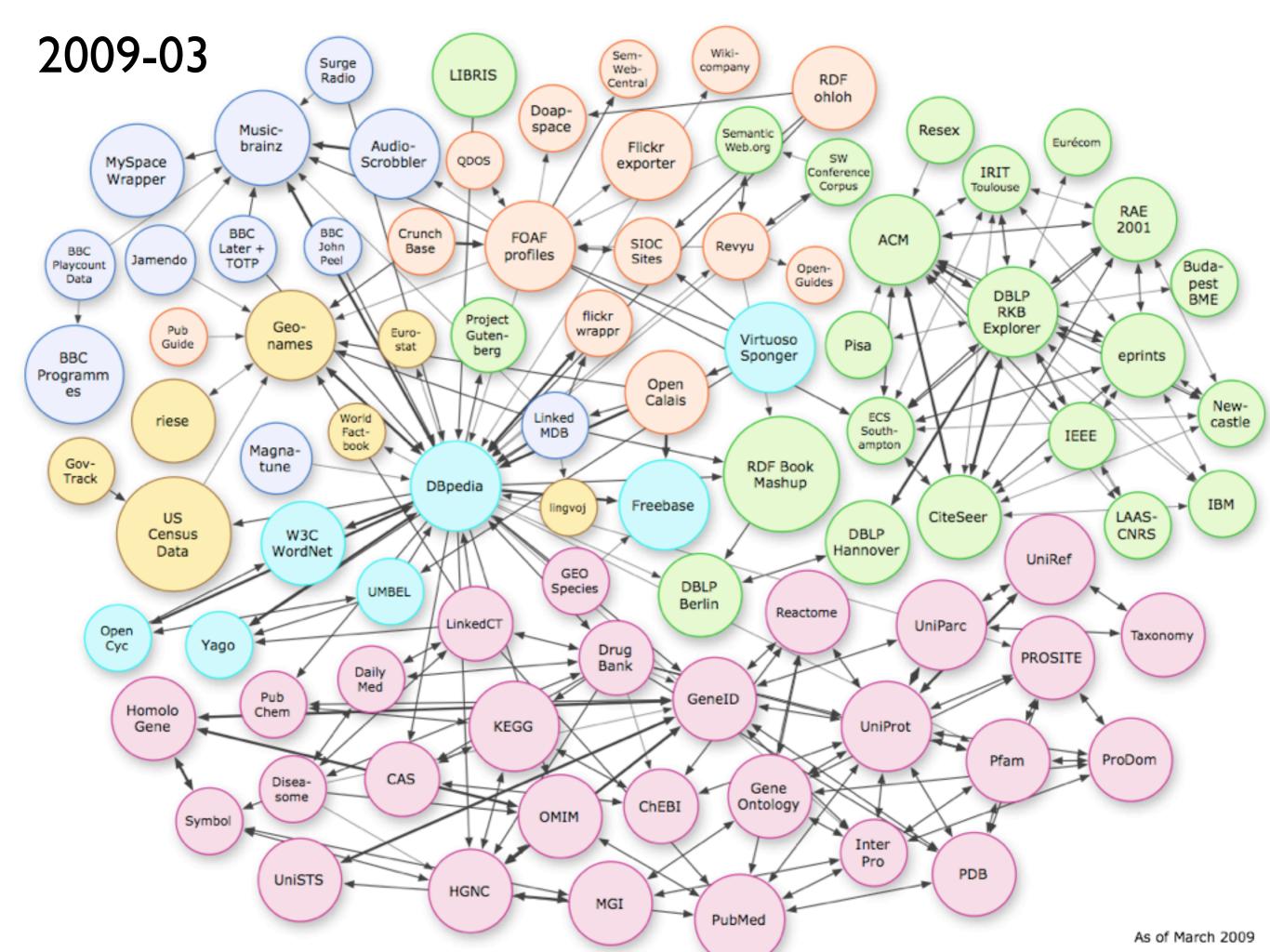
soon after, 'billion triples challenge'.

Linked Data

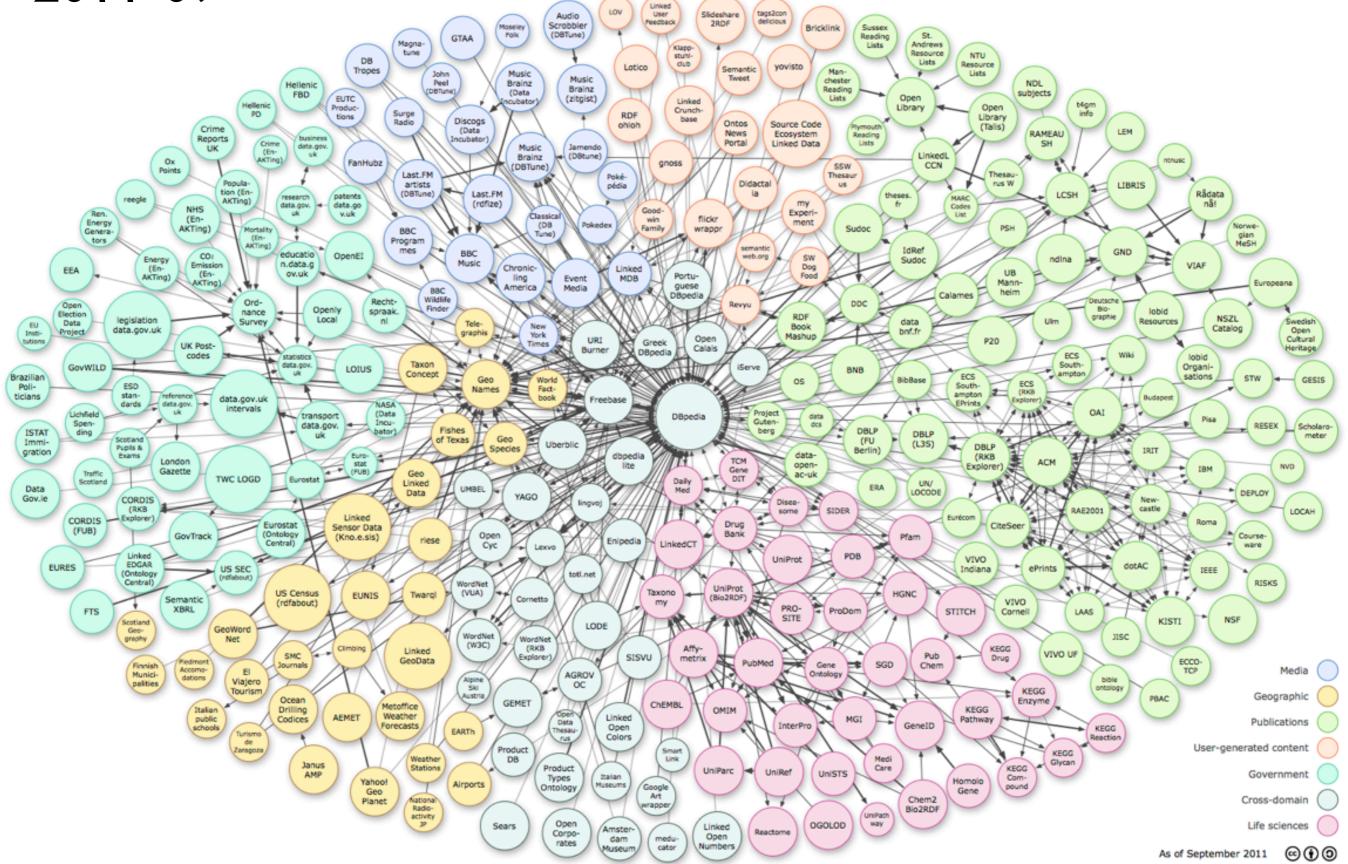
- The Linking Open Data community took this further, and quickly: publishing structured data from database-backed sites allows very rapid growth.
- LOD data was primarily in RDF/XML (rather than RDFa)
- The big shift: publishing structured data from databases
- LOD data typically assigned http: URIs to real-world things
- LOD structured data was rarely constrained by need to fit humanfacing HTML
- LOD structure typically authored by RDF experts using a diverse mix of several RDF vocabularies

2007-05





2011-09



The Big Shift

- LOD's explicit focus on publishing entire datasets boosted adoption massively.
- Assigning Web identifiers to real-world things made data merging and linking much easier than in earlier work.
- The cloud diagram inspired other outside the core RDF community; proof-of-concept conversions evolved into 'officially hosted' participation (e.g. LCSH.info & id.loc.gov)
- Technical publication details remained 'fiddly' (HTTP redirects, debates about URI structure) but an energetic community helped early adopters find their way.
- Buzz around "Web 2.0", APIs and open data fed this growth.

The Basic Idea: Links and claims

- In Linked RDF (both FOAF and Linked Data):
 - Pages (documents) express claims about the world.
 - Linked documents describe Linked entities.
 - This idea dates from the earlier days of the Web.
 - Understanding this gives an understanding of the entire design space we are exploring with Linked Data, schema.org and RDF-based languages.
 - Many technical issues are clearer when you realise there are two kinds of links involved...

June 2000:

The basic idea behind FOAF is simple: the Web is all about making connections between things. FOAF provides some basic machinery to help us tell the Web about the connections between the things that matter to us.

Thousands of people already do this on the Web by describing themselves and their lives on their home page. Using FOAF, you can help machines understand your home page, and through doing so, learn about the relationships that connect people, places and things described on the Web. FOAF uses <u>W3C</u>'s <u>RDF te</u>chnology to integrate information from your home page with that of your friends, and the friends of your friends, and their friends...

Dan lives in Zetland road, Bristol, UK with Libby and Craig. Dan's email address is danbri@w3.org. Libby's email address is libby.miller@
bris.ac.uk. Craig's is craig@netgates.co.uk. Dan and Libby <a href="mailto:work for an organisation called "ILRT" whose website is at http://ilrt.org/. Craig works for "Netgates", an organisation whose website is at http://www.netgates.co.uk/. Craig's wife Liz lives in Bristol with Kathleen. Kathleen and Liz also work at "Netgates". Damian lives in London. Martin knows Craig, Damian, Dan and Libby quite well. Martin lives in Bristol and has an email address of m.l.poulter@bristol.ac.uk. (etc...)

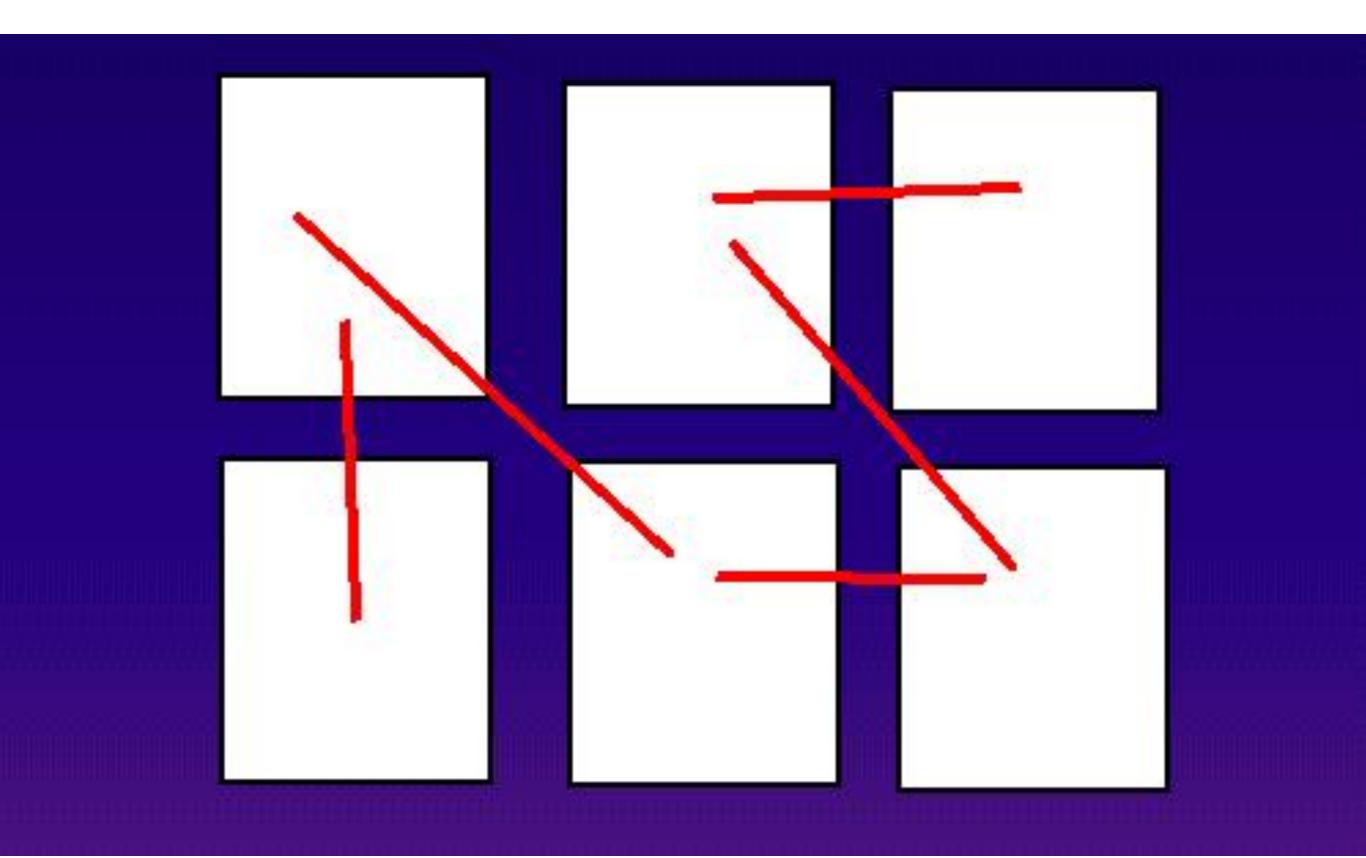
RDFWeb intro

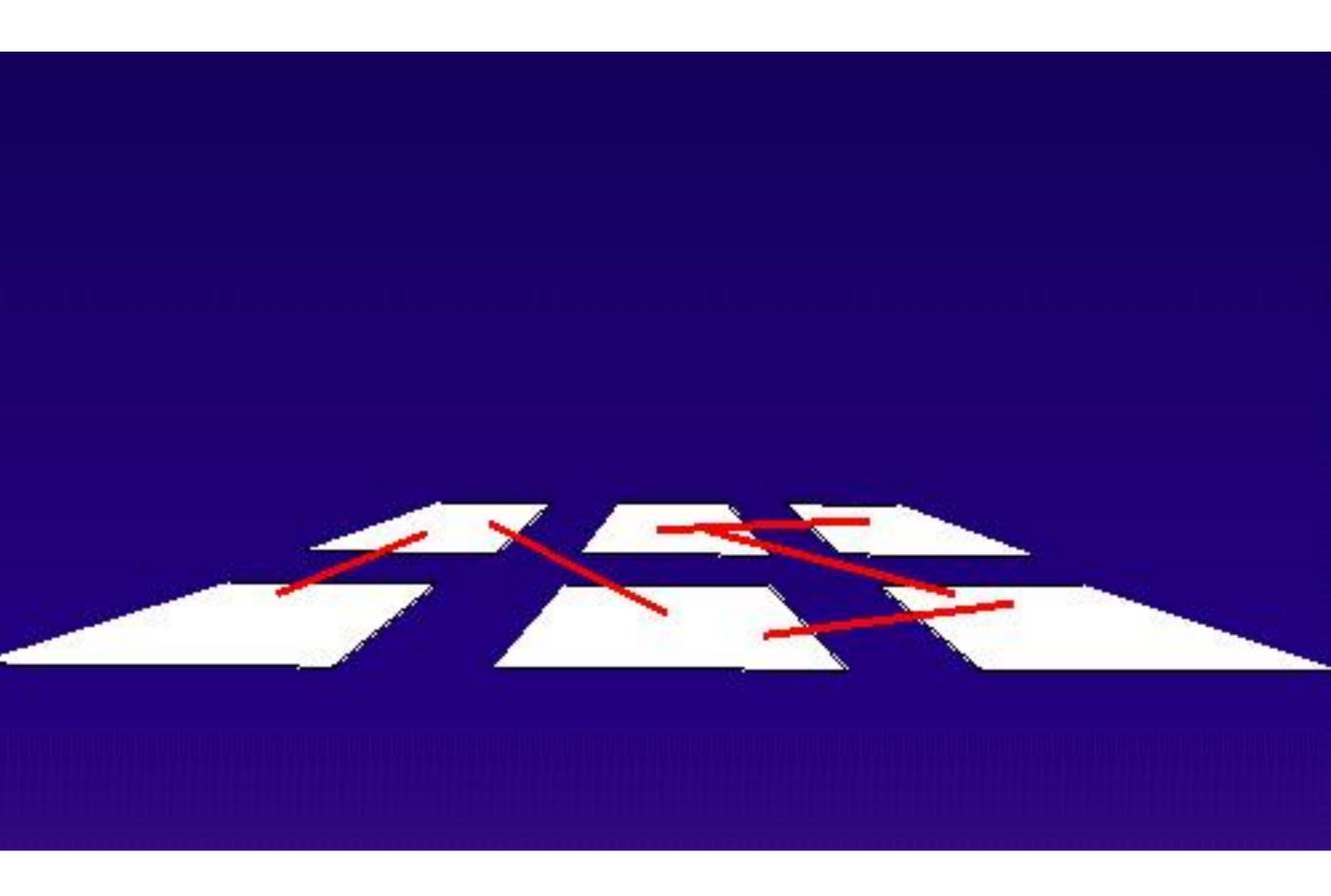
FOAF is best explained with an example. Consider a Web of inter-related home pages, each describing things of interest to a group of friends. Each new home page that appears on the Web tells the world something new, providing factoids and gossip that make the Web a mine of disconnected snippets of information. FOAF provides a way to make sense of all this. Here's an example, a fragment from the mostly-fictional FOAF database. First we list some facts, then describe how the FOAF system makes it possible to explore the Web learning such things.

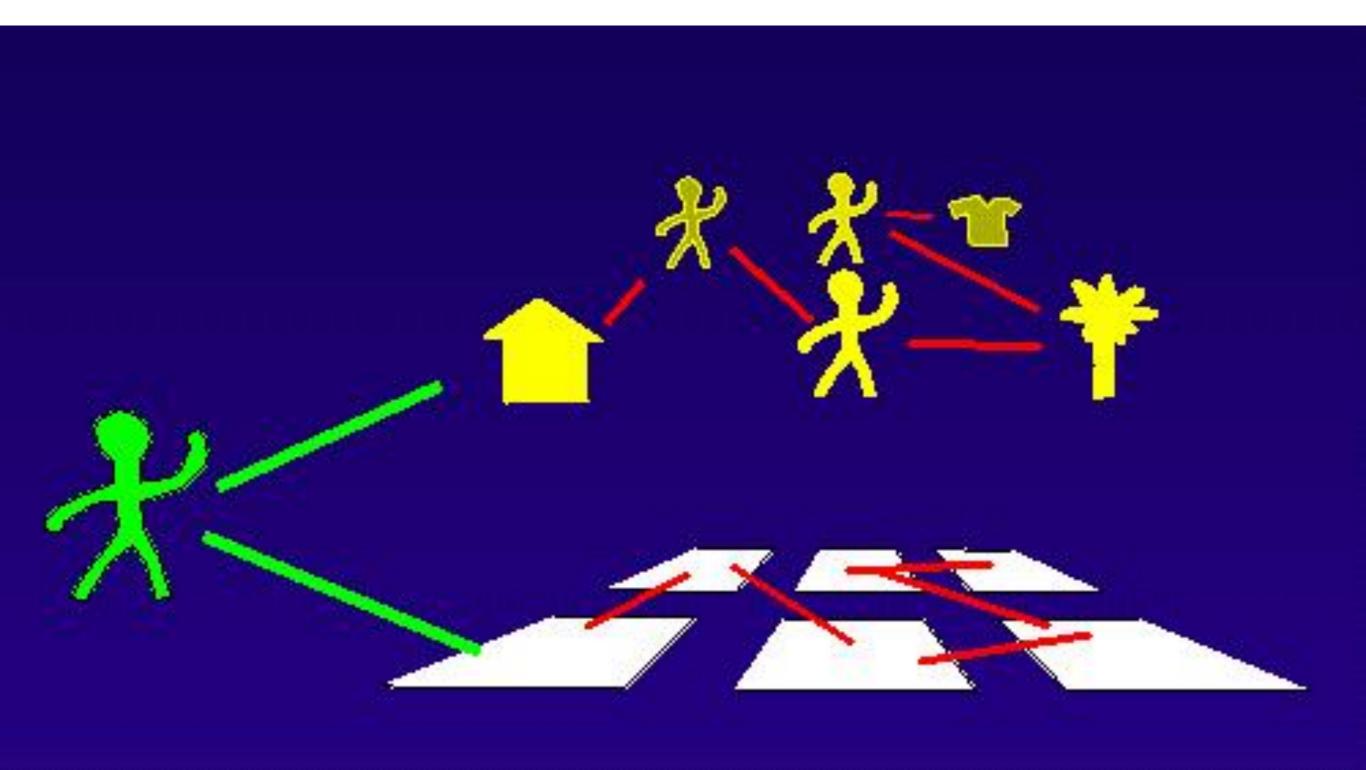
This kind of information is the sort of thing typically found on Web home pages. The extract shown here indicates how short, stylised factual sentences can be used to characterise a Web of relationships between people, places, organisations and documents. In real life, this information would be most likely be distributed across various Web pages created by the individuals listed. Very likely, their pages will link directly or indirectly to the home pages of countless other friends-of-friends.

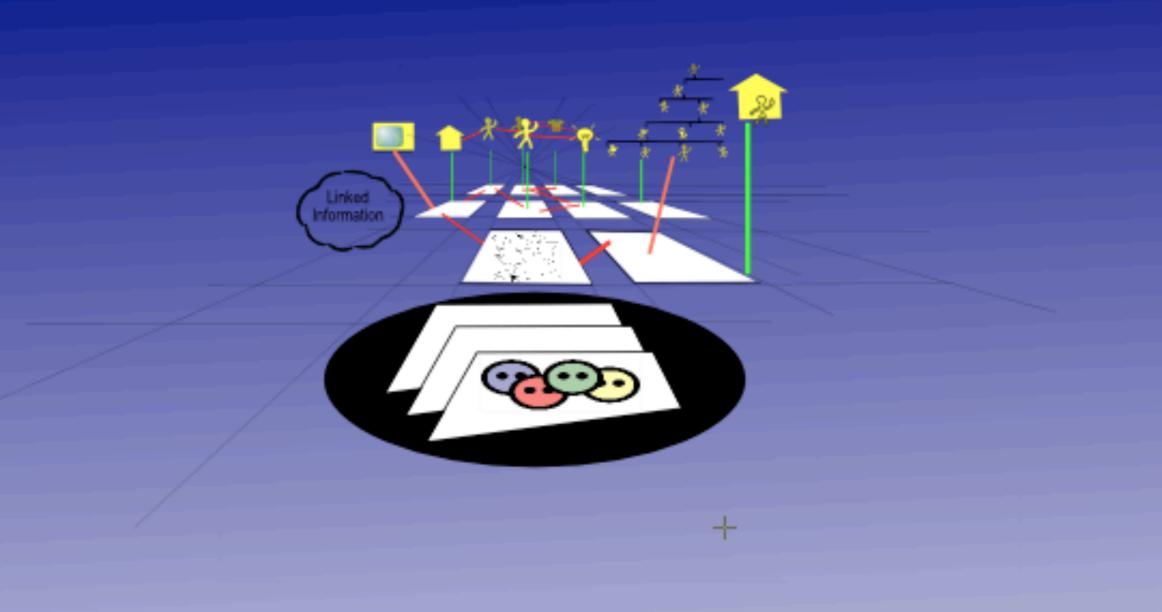
Claims

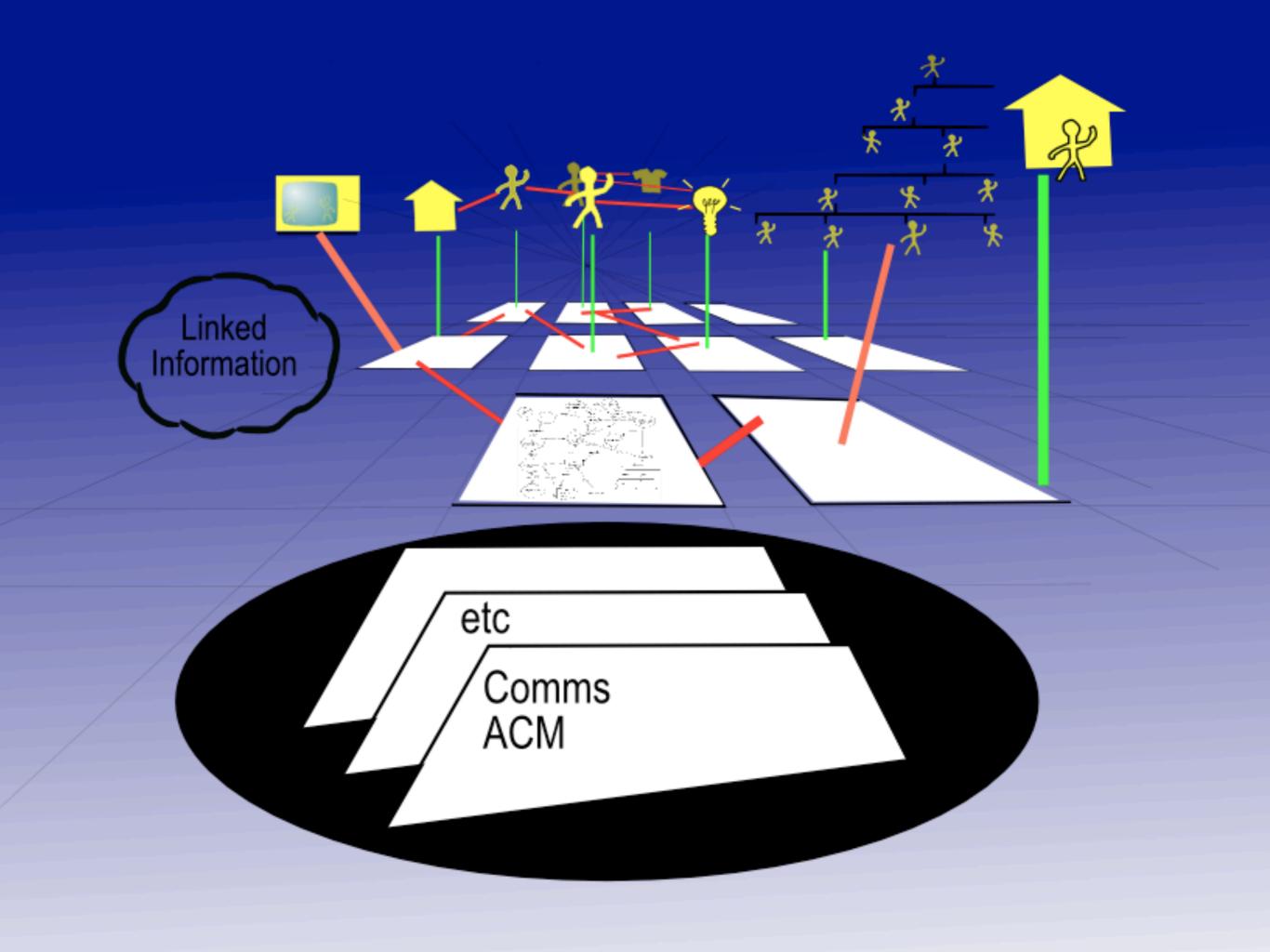
- Many pages are about some particular thing, e.g. a person or a movie or book.
- They describe various properties of that thing, often in terms of links to other things; these are *claims* that can be treated as structured data.
- RDFa, Microdata and Microformats explore this in HTML, Linked Data mostly in other formats.

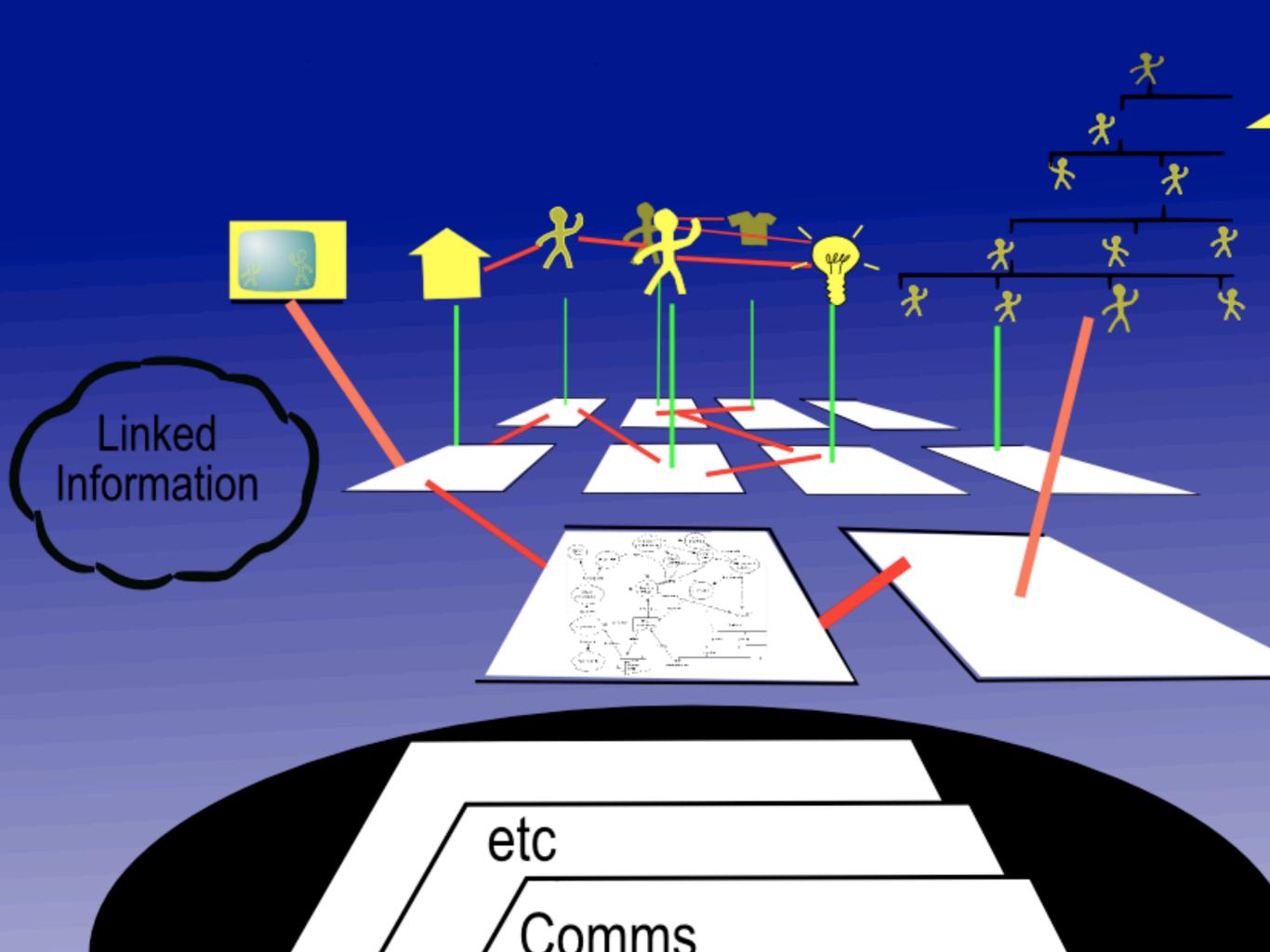


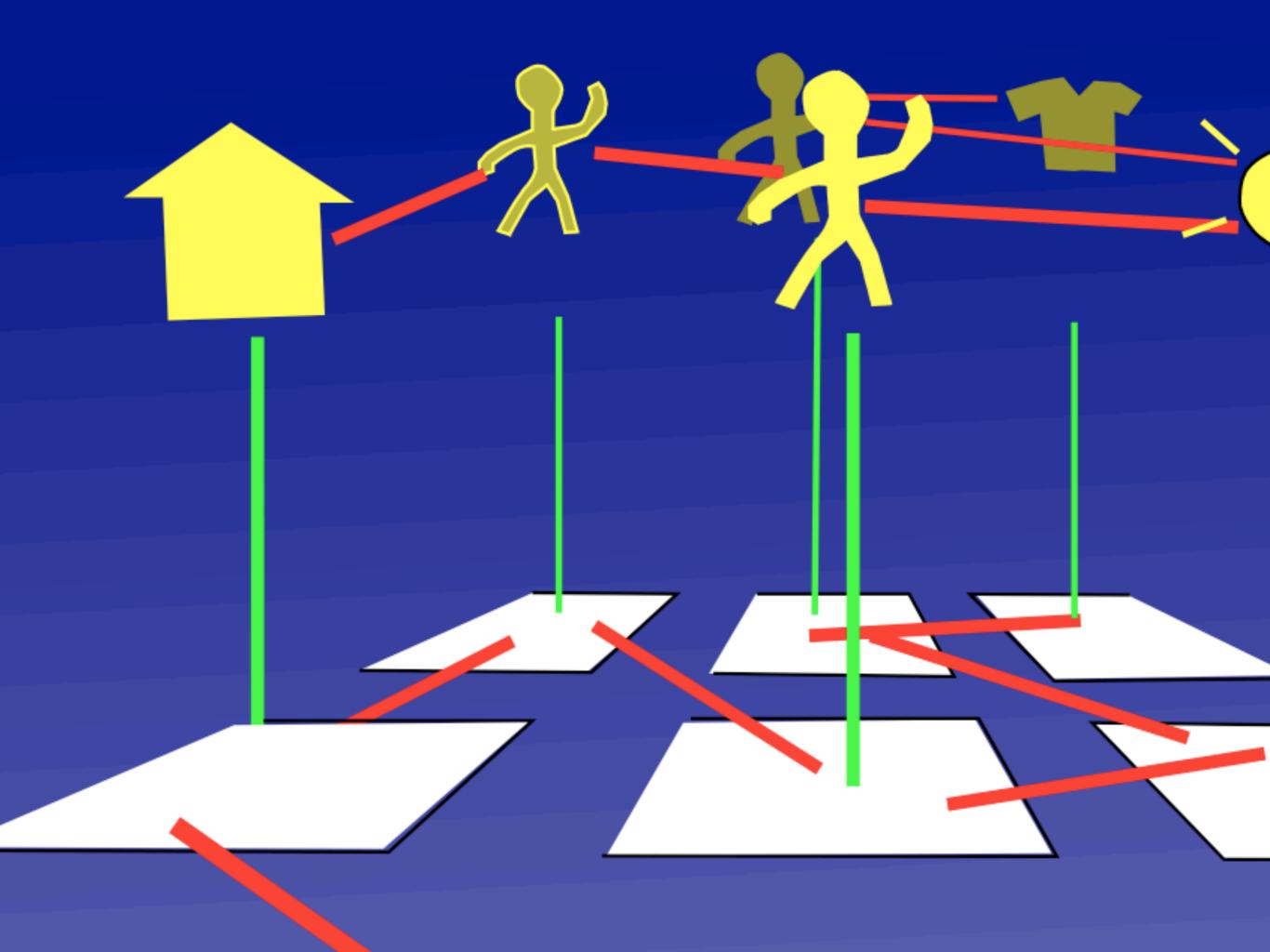


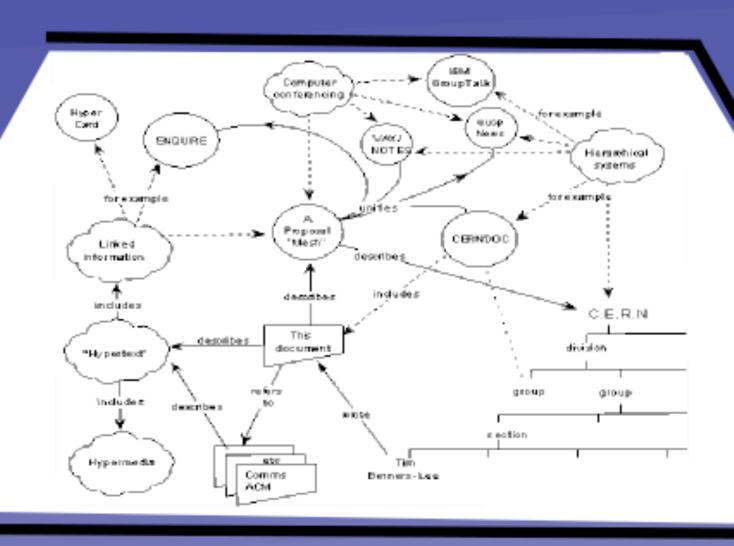


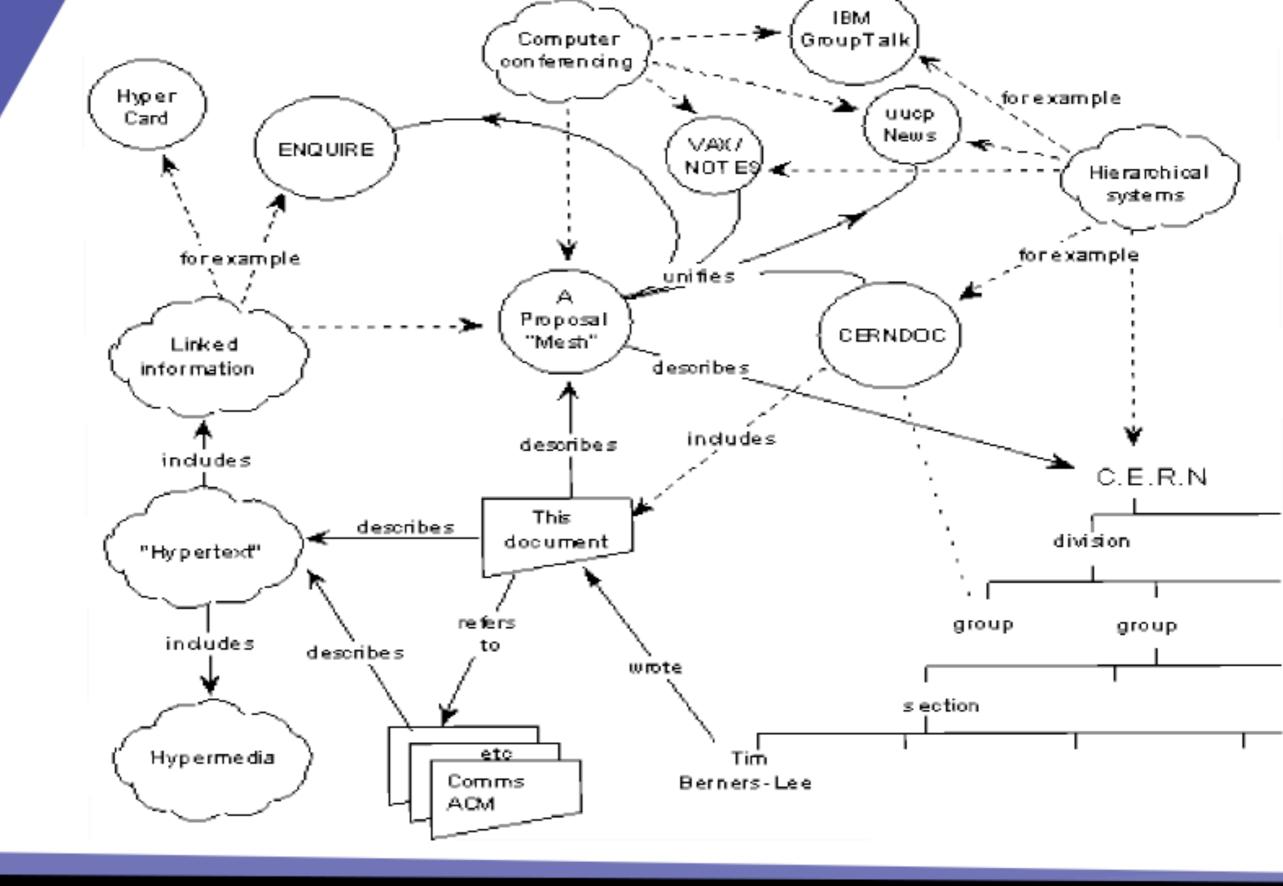


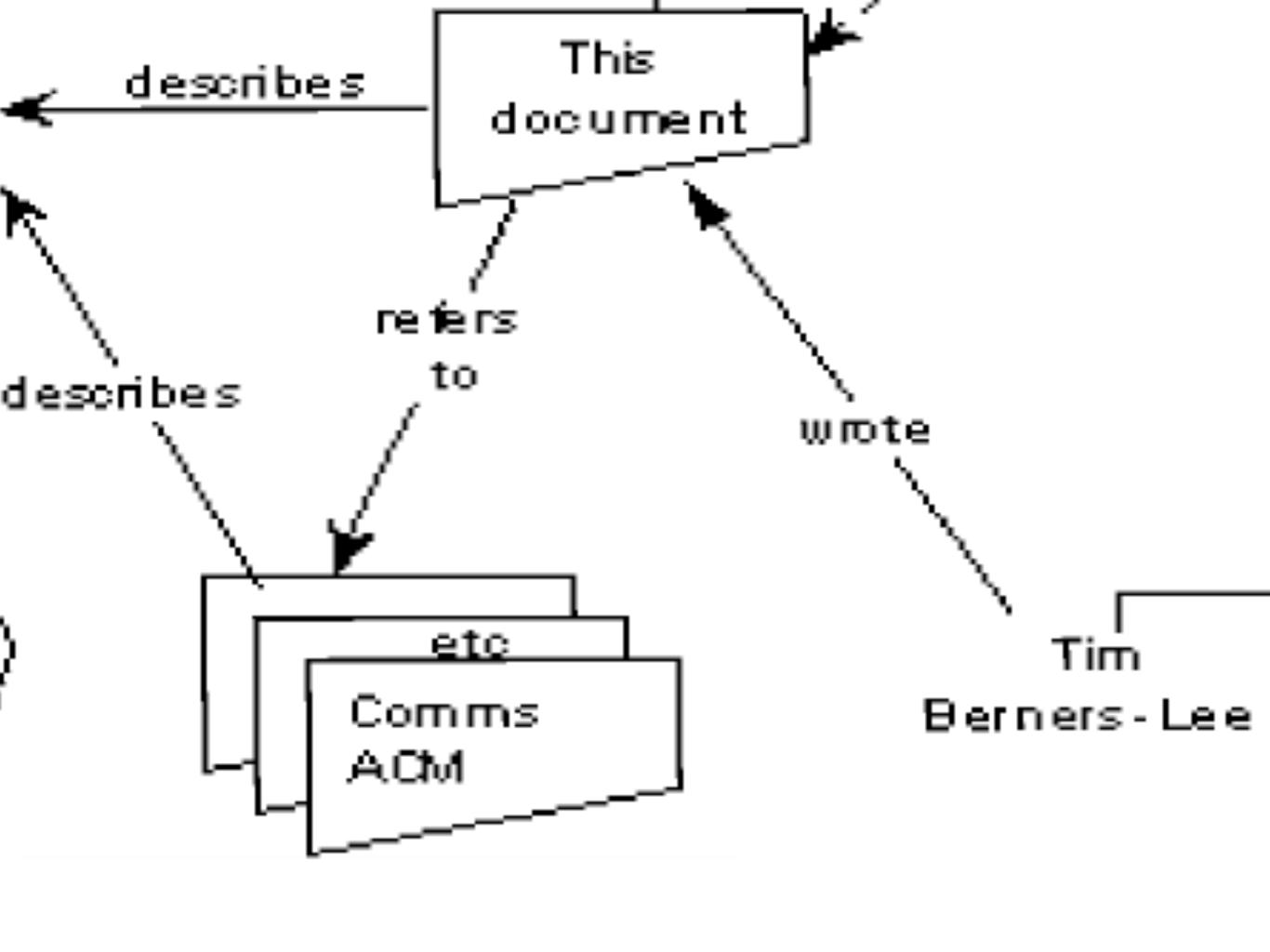


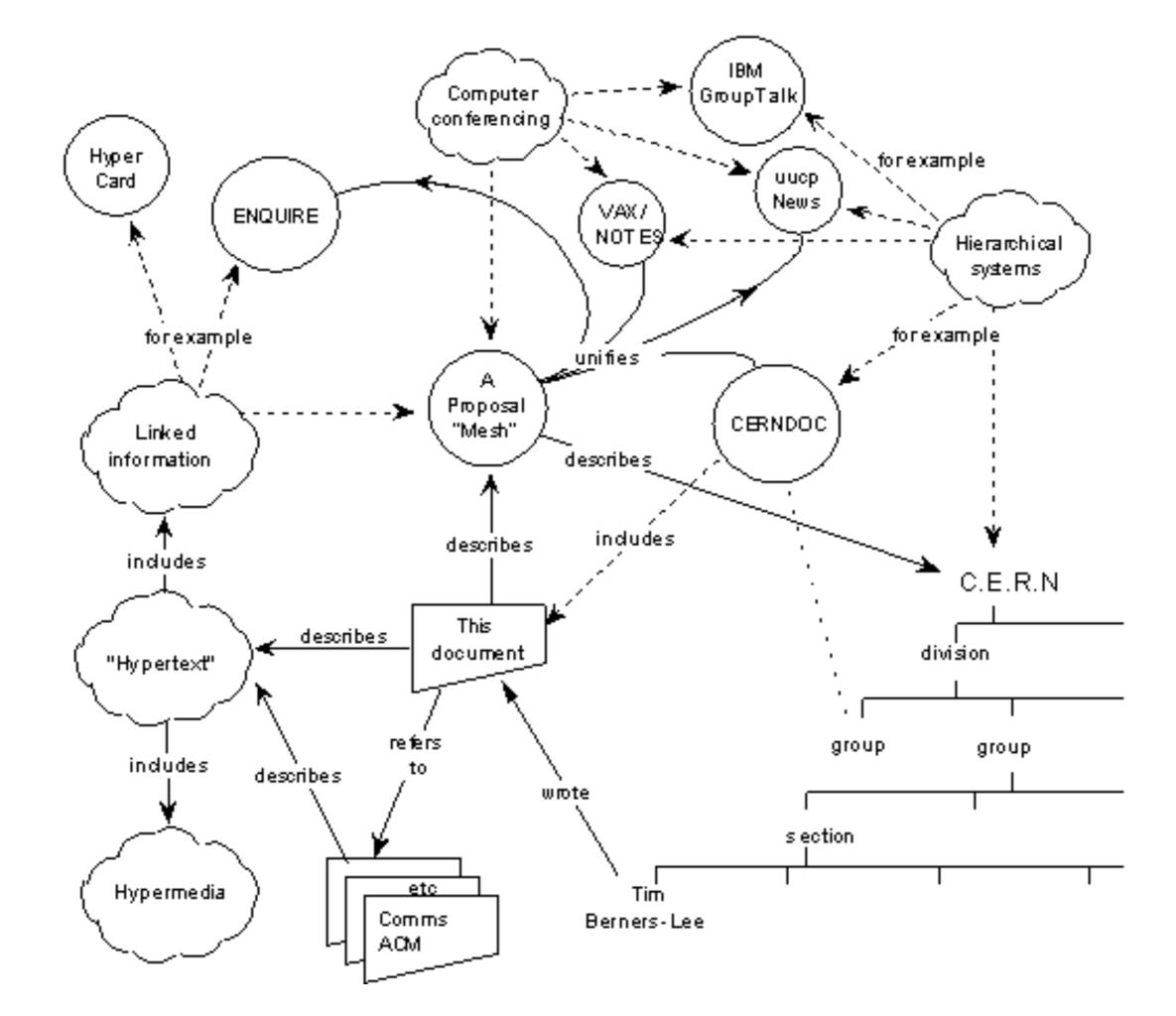


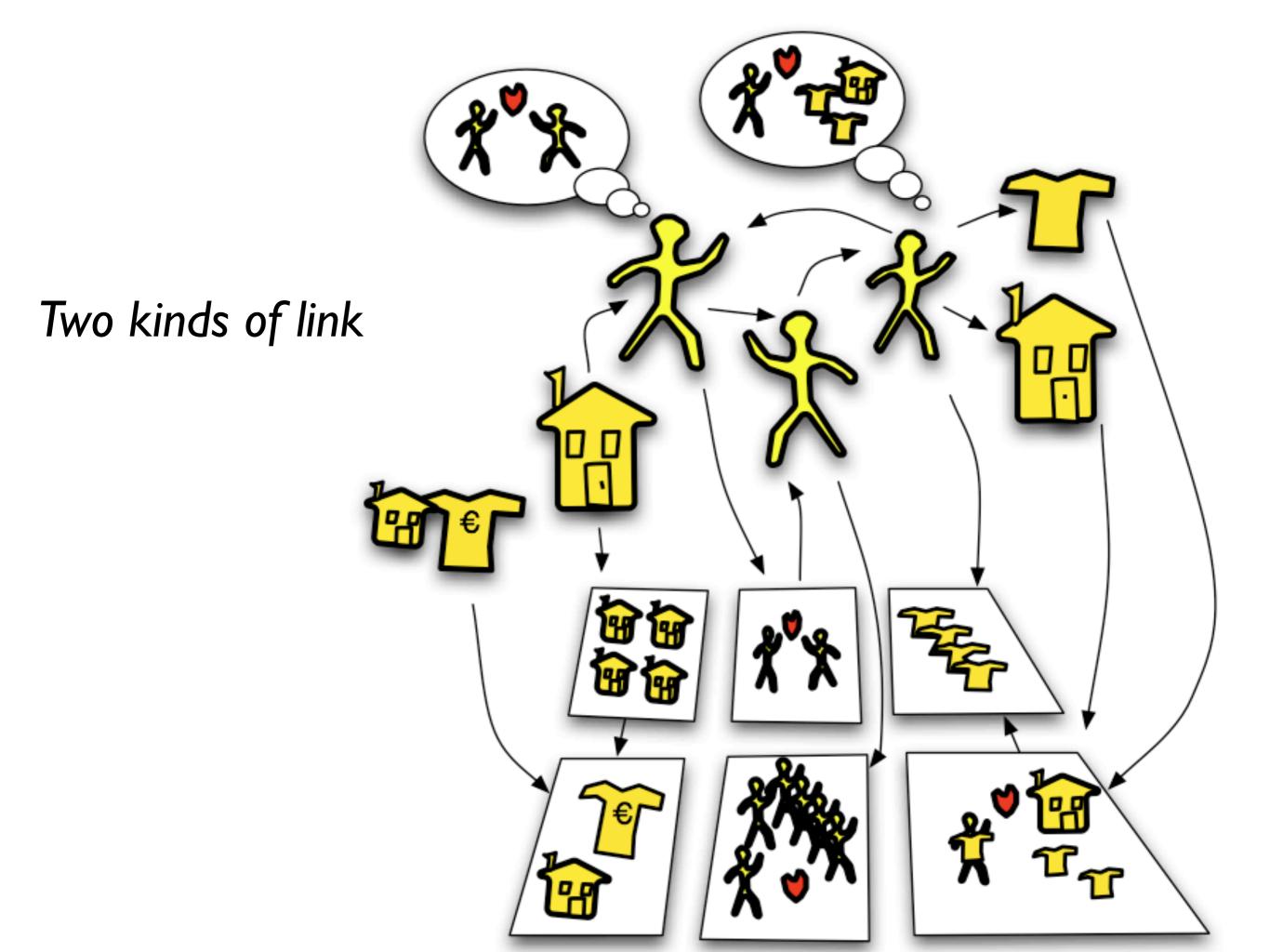












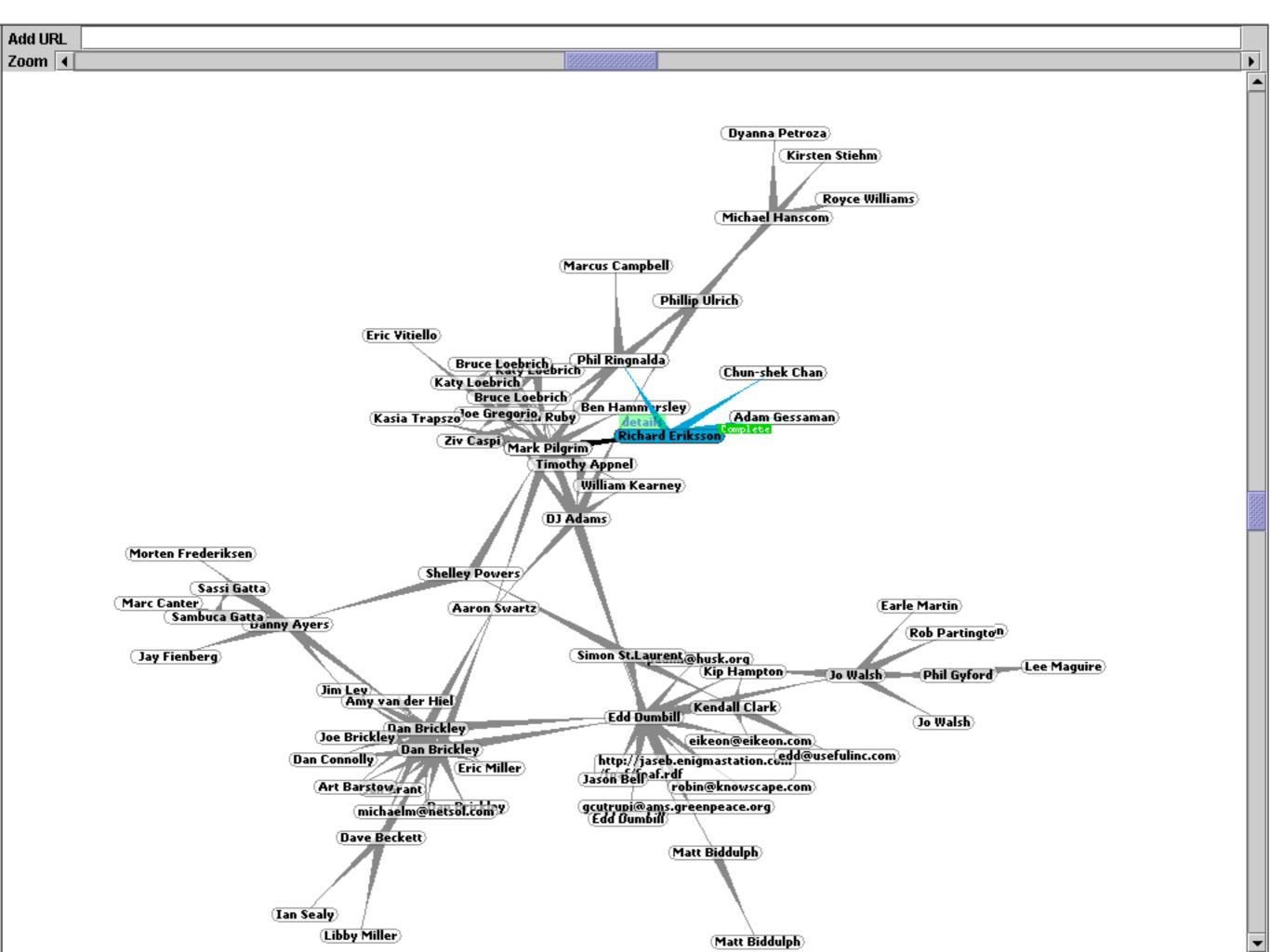
Links

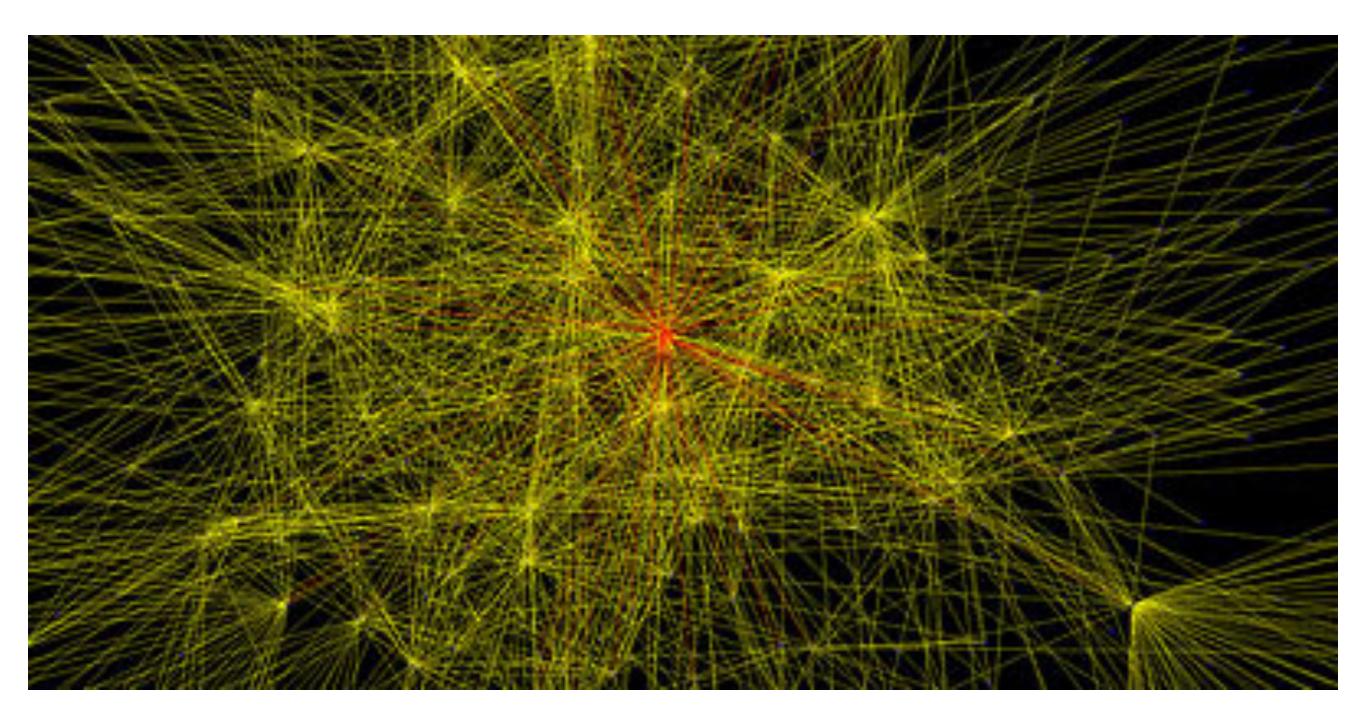
- Links between documents, links between things
 - As we compare Linked Data and schema.org, we'll see different tradeoffs being made.
 - Linked Data: an emphasis on richness of description and linkage, over human-readability and webmaster usability
 - Schema.org: simplifies expressivity and weaker linkage, as a tradeoff for greater webmaster/publisher usability
 - These are practical differences of emphasis and focus, rather than conflicts or different architectures
 - Both styles share the RDF data model, and can be blended together in several ways.

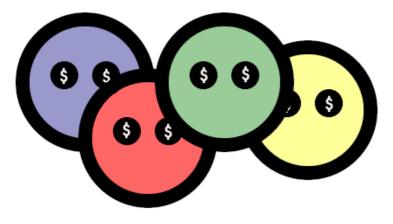
But why?

- Let's now take a few minutes to talk and think about incentives for publishing structured data.
- The incentives for publishing Linked RDF have matured over the years, and this has some effect on our technical choices...
- Who will be publishing, with what tooling, and for what reason?

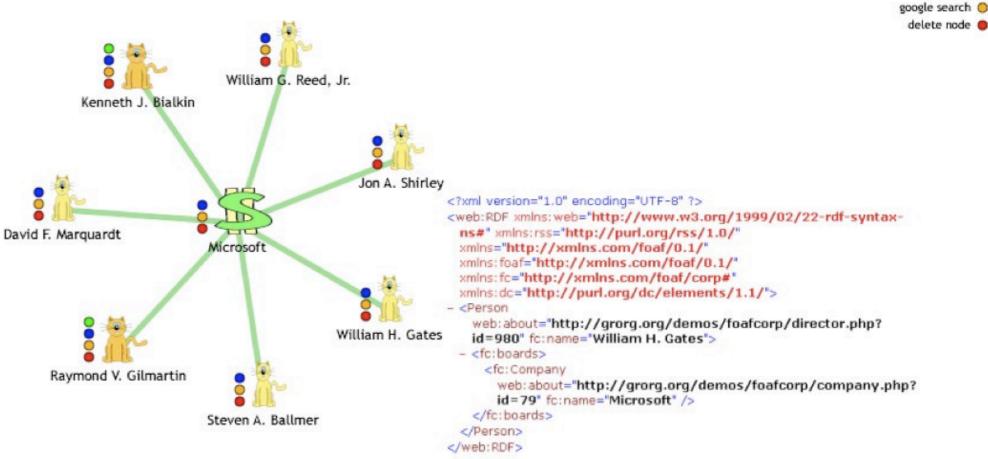
INCENTIVES



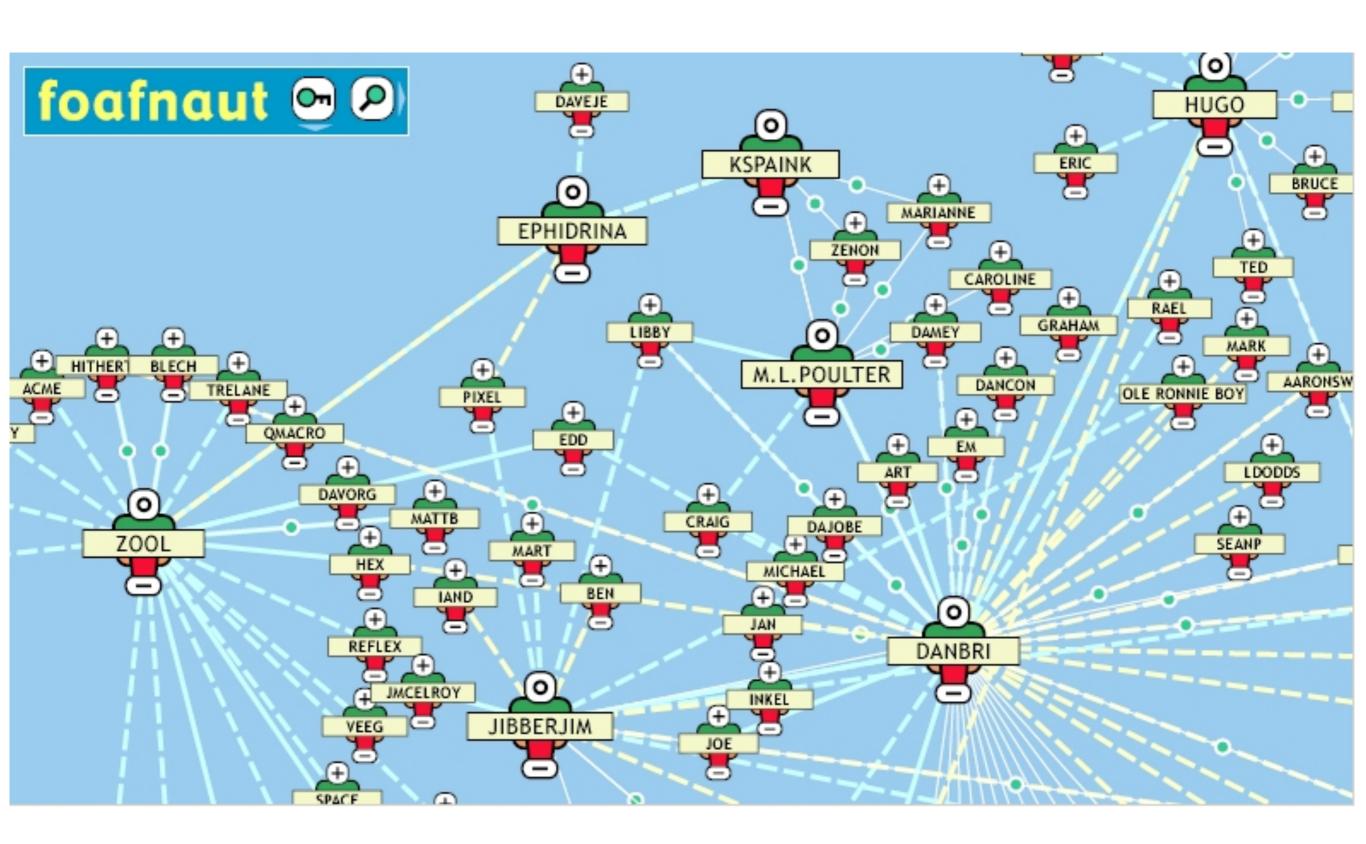




FOAFCorp



expand node o show rdf



Google Social Graph API



XFN & FOAF

Brad 1 - Friend David Mischaz Brod2)

and fitz com Live Janual (brad) bradfitz friend (bob) bob692)

Serious Linked Data



Incentives

- FOAF developer appeal 2000-7: fun demos; networks are intrinsically inspirational; general love of open data, APIs. Publisher appeal: it was easy and futuristic.
- Linked Data dev/publisher appeal 2007-2012: inspirational sense of an [open| linked] data 'movement' captured by the evocative and ever-growing cloud diagram.
- The data itself became a major asset; supporting technology was a means to an end.

Incentives: schema.org

- Schema.org: More (and better) Web traffic.
- A very mainstream message relevant to maintainers of millions of Web sites
- To sustain massive non-expert influx, some dramatic simplifications

Compare and Contrast

- Classic 'Linked Data' carefully distinguishes identifiers for things, and for pages about those things.
- Schema.org markup tends to leave the 'thing' identifier implicit, for simpler HTML.

Comparision (2.)

- Classic Linked Data: values rich, highly interlinked descriptions that use different independently created vocabularies.
- Schema.org: 300+ classes, 300+ properties in a single schema, addressing many more use cases without mixing vocabularies.