

# What we want with web-archives: will we win W8.0

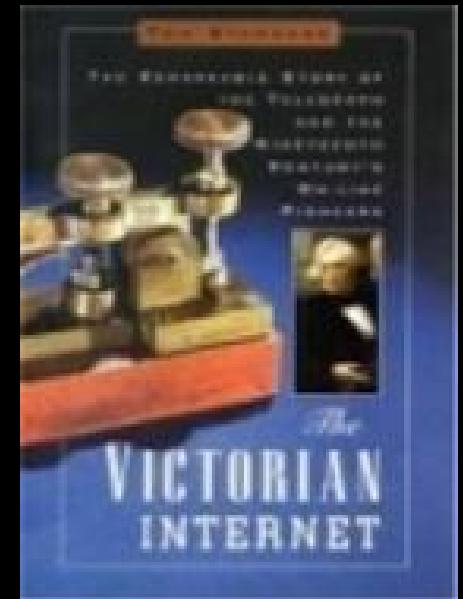
Kevin Ashley

ULCC Digital Archives Department

<http://dablog.ulcc.ac.uk/>

# Past histories

- Tom Standage – The Victorian Internet
- Not just what was said, but how



<http://vimeo.com/2312662>

# Thinking about use cases

- Not just document-centred
- Content
- Properties of content
- The web of data
- The web as data
- Stuff about the web as well as from the web

# Document-centred is useful

- For many academic uses, still central
- Sometimes content, sometimes presentation, sometimes both
- Timeslices or places over time:
- Brian Kelly's history of University of Bath homepage

# Content in aggregate

- Textual analysis
- Contrasting use of language
- Tracking spread of neologisms
- Word clouds

# Properties of content

- How quickly was PNG adopted ?
- Was takeup uniform in countries, types of site ?
- What did it replace ?
- What happened to XPM ?

# Searching the past

web archive - Google Search - Windows Internet Explorer

http://www.google.co.uk/search?hl=en&q=web+archive&meta=&aq=f&oq=

File Edit View Tools Help

Web Images Videos Maps News Shopping Mail more ▾ Sign in

**Google** web archive  Advanced Search Preferences

Search:  the web  pages from the UK

**Web** Show options... Results 1 - 10 of about 228,000,000 for [web archive](#) (0.20 seconds)

[Internet Archive](#)

**Web**, 150 billion pages ... Welcome to the **Archive**, See recent additions in RSS. The Internet **Archive**, a 501(c)(3) non-profit, is building a digital library ...  
[www.archive.org/](http://www.archive.org/) - [Cached](#) - [Similar](#)

<a href="#">Live Music Archive</a>	<a href="#">Web</a>
<a href="#">Texts</a>	<a href="#">Audio</a>
<a href="#">Moving Images</a>	<a href="#">Wayback Machine comes to life ...</a>
<a href="#">Grateful Dead</a>	<a href="#">Browse</a>

[More results from archive.org »](#)

[Internet Archive: Wayback Machine](#)

**Archive**-It allows institutions to build and preserve their own **web archive** of born digital content, through a user friendly **web** application, ...  
[web.archive.org/](http://web.archive.org/) - [Cached](#) - [Similar](#)

[Welcome to The Internet Archive Wayback Machine](#)

The Internet **Archive** Wayback Machine puts the history of the World Wide **Web** at your fingertips. The **Archive** contains over 100 terabytes and 10 billion **web** ...  
[web.archive.org/collections/web.html](http://web.archive.org/collections/web.html) - [Cached](#) - [Similar](#)  
[More results from web.archive.org »](#)

[UK Web Archive](#)

Creating an **archive** of culturally significant UK websites.  
[www.webarchive.org.uk/](http://www.webarchive.org.uk/) - [Cached](#) - [Similar](#)

[Web archiving - Wikipedia, the free encyclopedia](#)

2 Jun 2009 ... "Web Archive" redirects here. For the Safari format of the same name, ... **Web** archivists generally **archive** all types of **web** content ...  
[en.wikipedia.org/wiki/Web\\_archiving](http://en.wikipedia.org/wiki/Web_archiving) - [Cached](#) - [Similar](#)

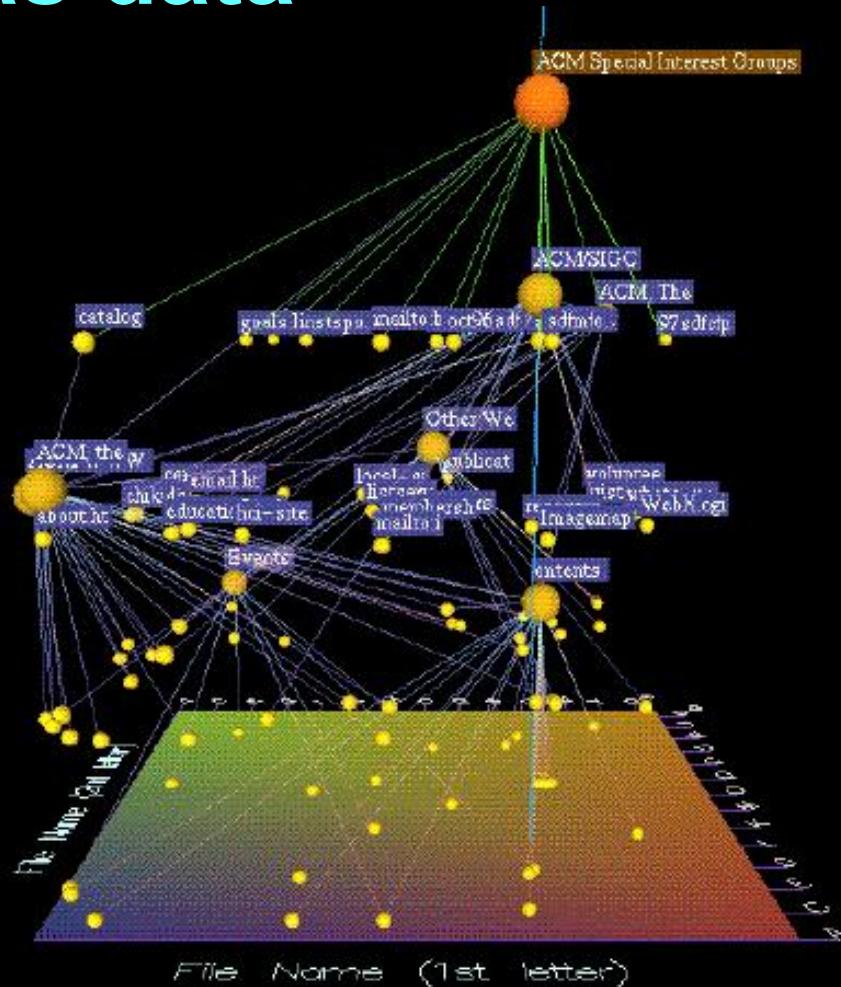
[Archives of Dead Web Pages: Wayback, Cache, and More](#)

27 Jan 2008 ... Lists and compares Internet current awareness services.  
[www.searchengineshowdown.com/others/archive.shtml](http://www.searchengineshowdown.com/others/archive.shtml) - [Cached](#) - [Similar](#)

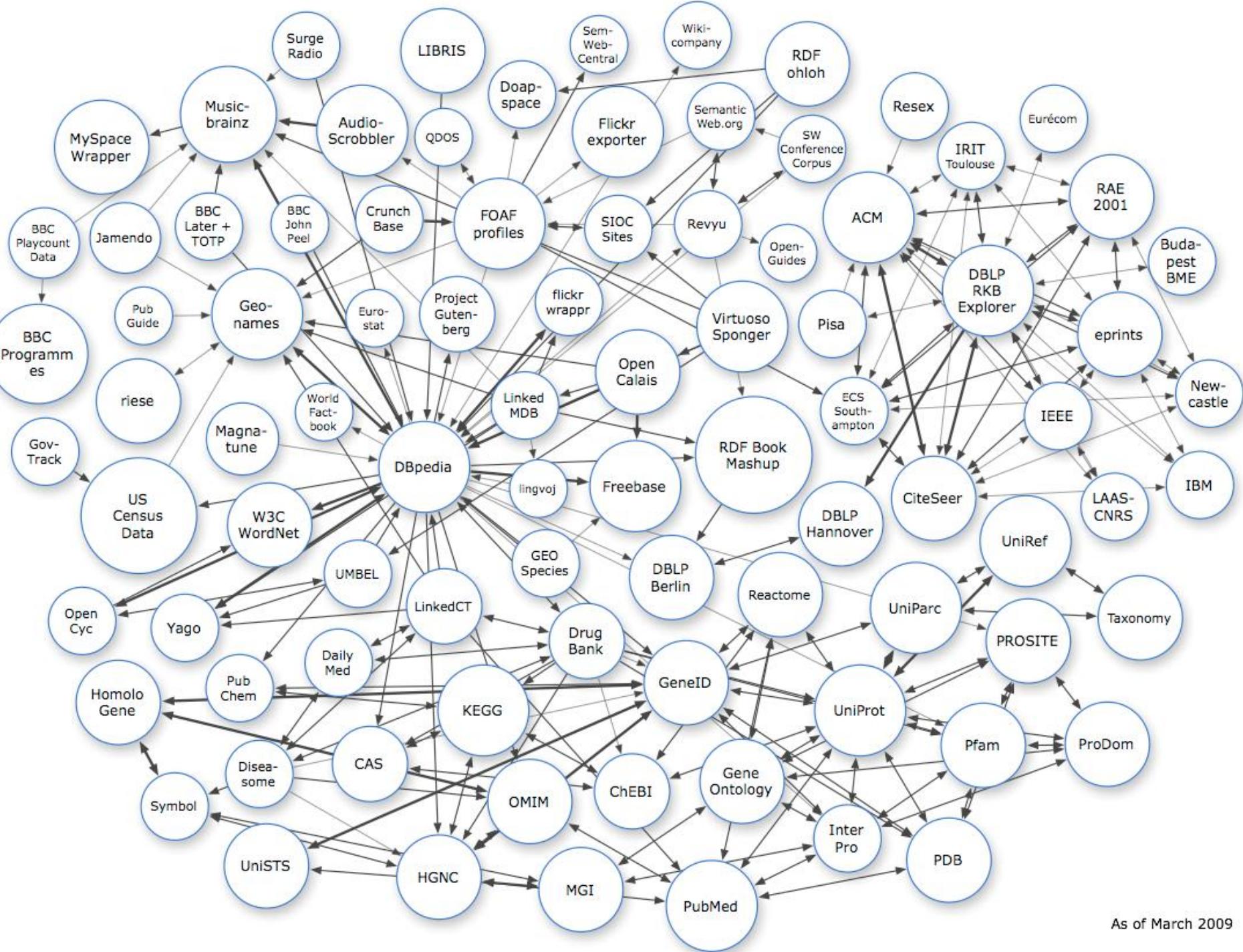
200 UK Government **Web Archive** | The National Archives 8

Start MyDocuments Microsoft PowerPoint - [...] [web archive - Google ...](#) Internet 100%

# The web as data



Hidekazu Shiozawa and Yutaka Matsushita – “Natto”



As of March 2009

# APIs that allow alternate views

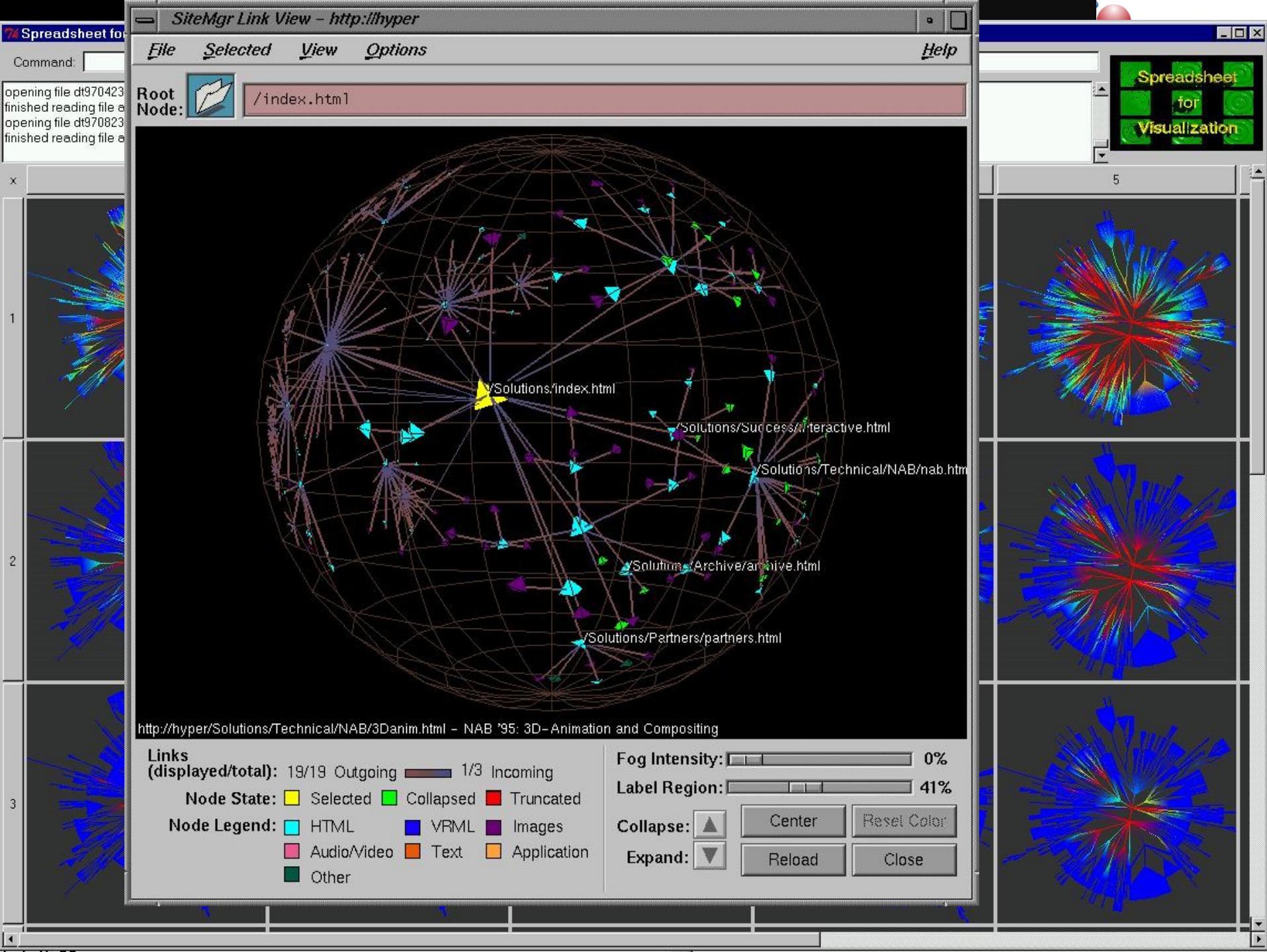
- Archives collect, protect and provide permanent references for content
- APIs allow many views and uses to emerge
- They permit intelligent intermediaries to do our work, or to assist
- Important as archive space fragments

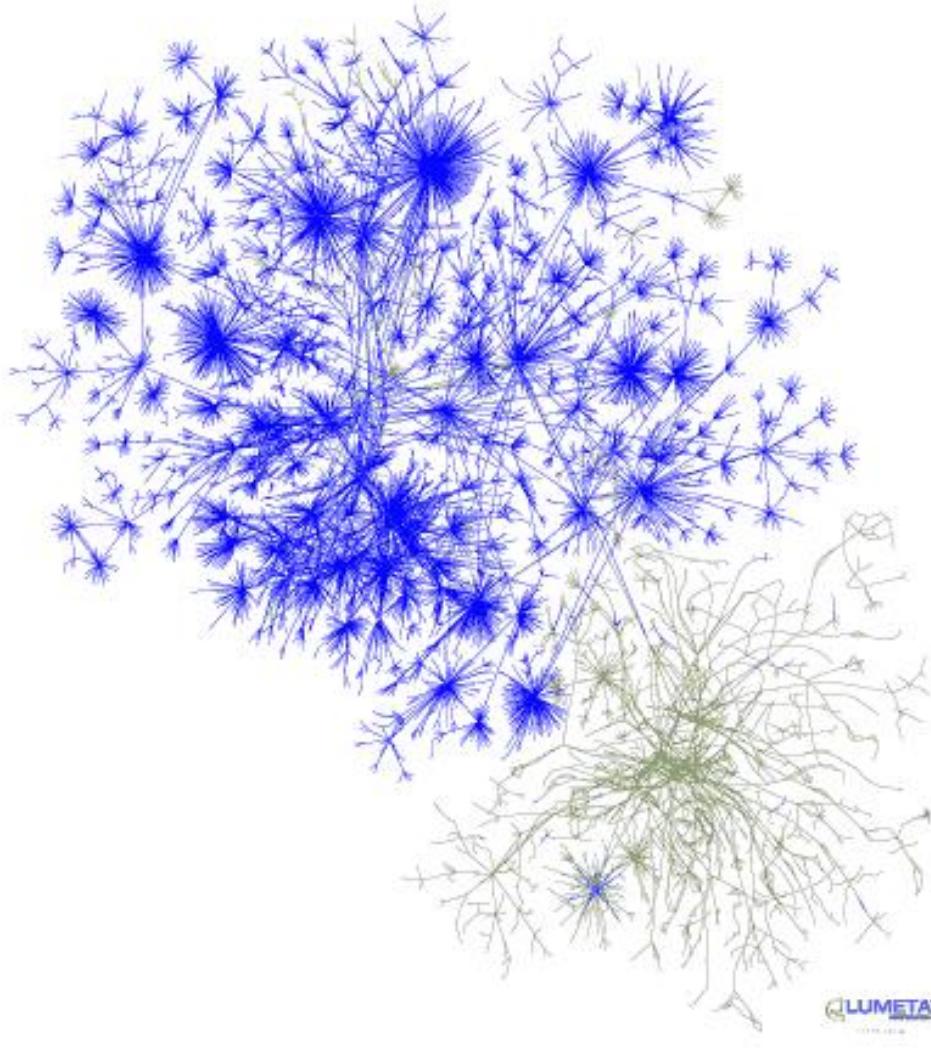
<http://taggalaxy.de/>

# Other stuff on or about the web

- Traditional media about the web
- Usage logs, server configs, server software
- Browsers, plugins, validators, ...







**Thanks to Martin Dodge's cyber-geography pages**