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
The web as data

Hidekazu Shiozawa and Yutaka Matsushita – “Natto”
The web of data

• Linked data: "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web"
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, …
Kevin Ashley: http://dablog.ulcc.ac.uk/
Thanks to Martin Dodge’s cyber-geography pages