Towards A Macroscope For UK Web History

Andrew Jackson
UK Web Archive Technical Lead
The UK Web Archive

• Three collections:
  – Selective Archive (since 2004)
  – Legal Deposit Archive (since 2013)
  – JISC Historical Archive (1996-2013)

• Statistics:
  – Over six billion resources
  – Over 100TB compressed data

• Goals:
  – Preserve UK web history
  – Support access
  – Enable research
The Problem
The Solution?
The Macroscope

“Macroscopes provide a 'vision of the whole,' helping us 'synthesize' the related elements and detect patterns, trends, and outliers while granting access to myriad details.”

– Plug-and-Play Macroscopes by Katy Börner
– Big Data, Small Data and Meaning by Tim Hitchcock
The Demo

http://www.webarchive.org.uk/shine
The Future

• Scale & Sustainability:
  – 1996-2010 – 2 billion resources
  – 2011-2013 – 2 billion more in *two years*…
  – 2013 onwards – >2 billion *per year*…

• Features & Usability:
  – Better full text search result quality, and more *curation*
  – Richer deconstruction of trends
  – Contextualisation of individual URLs and hosts
  – Improve network analysis capability
  – Understand lifespan and lifecycle of URLs