Applying the Principles of the UK Government Web Archive to Social Media

Tom Storrar
Head of Web Continuity
Outline

• Brief overview of the UK Government Web Archive

• Evolution of scope: websites and social media

• Selection

• Scope of capture

• Archiving workflow

• Access
Collection

define and collect the future record, transparently and seamlessly

Preservation

innovate to ensure optimal physical and digital preservation

Access

provide a record that is more open, inclusive and used than ever before

The National Archives
Applying records management processes to the open government record 1996… 2014

The UK Government Web Archive
Selection

- Initially, a pilot project was devised with supplier, Internet Memory Foundation, to develop capability to capture more complex social media.
  - Interviews with government departments: which services do they use the most?
  - Analysis of social media services and their archiveability (i.e. YouTube, Twitter, Facebook, Flickr)
  - Decided to focus on Twitter and YouTube based on government’s and end users’ needs and technical feasibility

- The National Archives’ OSP27 document has gone through several iterations between 2007 and now:

- The 24 ministerial departments, plus the 22 non-ministerial departments are in scope automatically for websites and for social media archiving.
- Other organisations – e.g. Environment Agency – and event-driven social media – e.g. London 2012, YouDecide2014 (Scotland Independence Referendum)
Scope of capture

• Within UK legislation, we have to make sure that we only capture content created by government departments. This means:
  – no retweets (Twitter) or comments (YouTube) can be captured, making it a one-way conversation. When a non-government user Tweets at a department we do get replies to @s, but not the original Tweet. For example:

  ![DCMS Tweet]

  – …but the majority can be viewed in isolation, e.g.

  ![DCMS Tweet]

  – we must be adaptive to changes in the status of organisations. For example, a government body *may* cease to be a government body after a certain time so we must make sure that we only capture content created during the period that it is a government body.
Video Archiving

Collect
- Unlimited number of Videos
- Video metadata
- User comments

Preserve
- Store video is ARC files, in HDFS, with metadata in DB

Access
- Video search
- Reuse enabled
- UKGWA content discovery

Quality Assurance
- QA
- UI
Questions?

Contact us: Web.Archive@nationalarchives.gsi.gov.uk
Visit: nationalarchives.gov.uk/webarchive