Web Archiving Roundtable minutes

Wednesday, August 13, 2014

Marriott Wardman Park, Washington, D.C.

* **Welcome from Tessa**
  + Change in agenda - Speakers first
  + Intro to new committee members - Kate Stratton (new Co-Chair), Rachel Taketa (new Social Media Manager)
* **Jimmy Lin, Associate Professor at iSchool @ University of Maryland**
  + Twitter experience: creating analytic infrastructure for data mining, developing data products to surface relevant content to users
  + Hard to build tools for non-existent users - need deep collaboration with users (archivists librarian, journalists, researchers) and technologists
  + Needs to go beyond browsing and searching
  + Using open-source big data tools (Hadoop HDFS & MapReduce; Apache HBase) been building WarcBase, an open-source platform for web archives (warcbase.org)
  + Tools and services built out from WarcBase - Topic Model Explorer, WebGraph Explorer
  + Question about visualizing inconsistencies between sites (changed crawl scopes, or removed content, etc)
  + Question about optimal service model for these tools/services
  + Question about use - for the research community or for "us"
* **Jason Scott, Founder, Archive Team**
  + Started in 2009, to go after websites that were going down for the future
  + Grabbed GeoCiities
  + People said "who cares, that time is over"
  + "the bro-grammer-journalism complex defined our culture for us"
  + Archive Team Warrior - creating an army - likened to a "low orbiting ion cannon"
  + Disenfranchise, Demean, Delete -
  + Facebook is essentially run by vodka-drunk children
  + 3 virtues of Archive Team - paranoia, anger, and kleptomania
  + Ethics of web archiving
* **Nicholas Taylor, 2013 NDSA Web Archiving survey results overview**
  + 2013 NDSA Web Archiving survey results
  + Imagined as bi-annual survey - first in 2011, latest in 2013
  + Respondent characteristics - Universities still making up most
  + Affiliations - WART wins 45%
  + Most have fractional staff dedicate to web archives
  + Signs of increasing maturity: more active programs
  + Strong perceptions of progress in their programs
  + Focus of programs: changes in number doing only other sites or their own sites
  + Concern about social media, databases, and video
  + Untapped interest in collaboration
  + Web-archiving-as service still most popular option, most don't transfer data from vendor
  + Increase in tools supporting WARC
  + Decrease in every descriptive option except new ones
  + Most are not looking for permissions
  + more conditional handling of robots
  + Most lack social media archiving policies (75%)
  + Most rely on other organizations policies for informing their own policies
  + What collabos should we be doing? thinks WART is critical
* **Anna Perricci & John Bence,**

[SAA Web Archiving Roundtable Education Needs Assessment Survey Results](http://www.slideshare.net/annaperricci/saa-web-archiving-roundtable-education-needs-assessment)

* **Adjournment and a big THANK YOU from Tessa Fallon, outgoing chair**