

Citizen Cyber Archives: Sustaining Archives through Communities

Mark Hedges, Richard Palmer
Centre for e-Research, King's College London, UK

Archives, physical and digital

- Physical archives (paper etc.) ...
 - Physical access may be difficult
 - Hard (and time-consuming) to find things
 - Dangers to sustainability (disposal, fire)
- ... to digital archives
 - Easier to access/locate (maybe)
 - But not the end of the matter

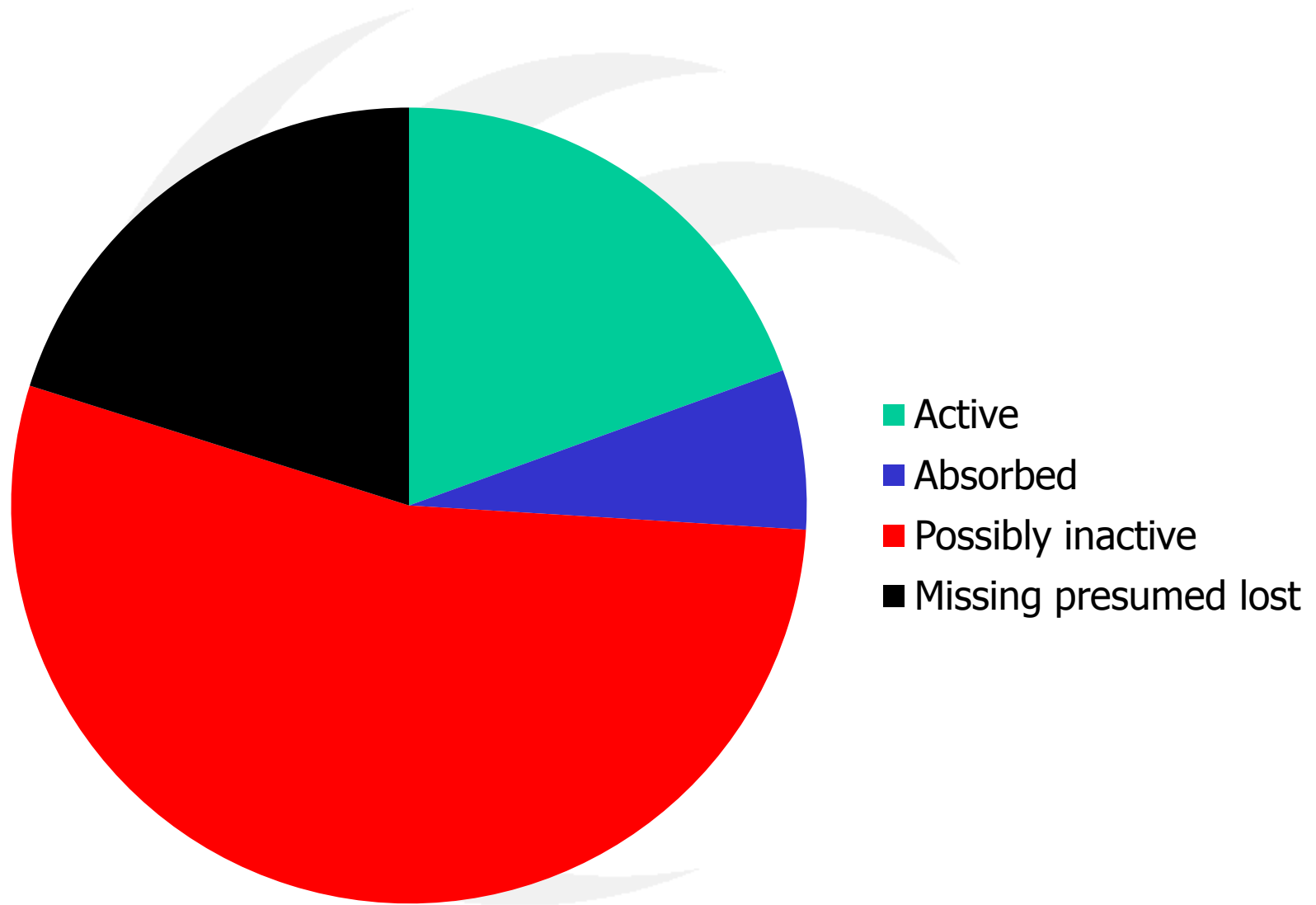
Sustainability: an example

- NOF-Digitise Programme
- Funded by “New Opportunities Fund”, part of UK National Lottery
- Ran from 1999 to 2004
- Awarded £50,000,000 to ~155 projects
- Digitise resources and make them available

Whatever happened to ...?

- 30: Website exists and seems to have been enhanced since the launch
- 10: Absorbed into larger website
- 83: Website exists and seems not to have changed since launch
- 31: No URL available, or URL doesn't seem to contain the digitised material
- <http://web.me.com/xcia0069/nof.html>

[Thanks to Alastair Dunning of JISC - data last updated 22 August 2009]



Citizen Cyberscience Summit 2010, London – 3rd September 2010

Many ways an archive can die

- The server vanishes
- Available but unchanged
- Technical obsolescence
- Content obsolescence
- Inadequate metadata (poor discovery)
- Hidden archives
- Isolation (from the web of data)

Involve the community?

- Many archives (could) have communities around them
- Exploit the interest of specific groups for specific archives

facebook

- Tagging?

flickr

You Tube

- Simple, but ...
- Too libertarian, open to misuse
- Not appropriate for more formal archives of heritage material
- <http://www.steve.museum/>

Volunteer Thinking

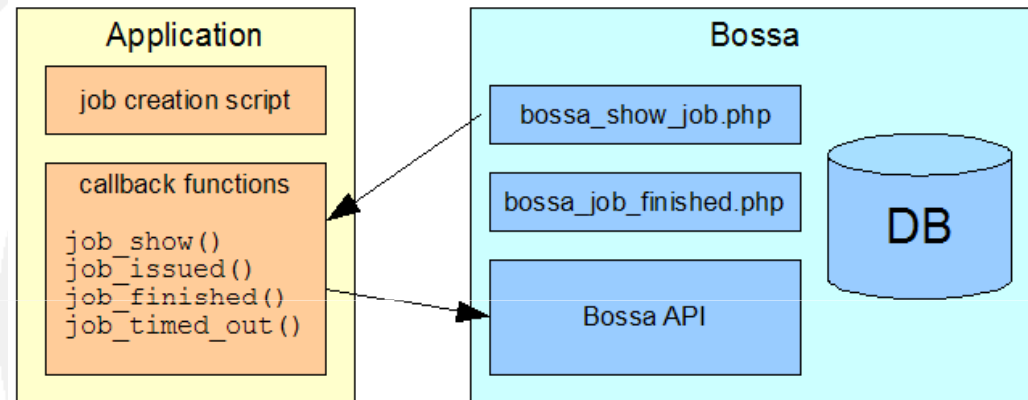
- Divide work into small, individual tasks, assigned and carried out via the web
- Difficulty of tasks varies:
 - Performed by general public, without specialised knowledge (but maybe training)
 - Some improved by specialised knowledge
 - Some require specialised knowledge
- Some (not all) tasks dull and repetitive

Accuracy and Reliability

- Task replication (consensus)
- Calibration of volunteers' performance
- Quantification of performance
- Better performers can do harder tasks
- (Partial) validation by experts

Bossa

- Provides software framework



- Allows us to:
 - create and assign tasks,
 - evaluate volunteers
 - define policies for task replication and assignment (taking evaluation into account)

Archives@Home

- Apply methodology to digital archives and cultural heritage collections
- Sustain archives (help them survive)
- Enhance archives (make them more useful)
- Involve specific communities
- “Adoption” of archives



TELDAP

- Very rich archive of material relating to Chinese and Taiwanese cultural heritage
- Does not have high visibility worldwide
- Make more globally accessible:



- Metadata enhancement
- Metadata translation
- Creation of fine-grained item metadata
- Thematic metadata and themed meta-collections



Great Ormond Street Hospital



← GOS Hospital, 1875

Dr. West.		358	
Name	Ellie Pope.	Sex	Age 84 1/2 yrs
Address	31 Roaston Pl. Dorset Sq.		
Admitted	5- Jan. 23. 1864.	DISEASE.	Date of Attack
Delivered	Apr. 4. 1865	Chorea - severe, caused, most prob. by affection of heart, except irregularly, but recovery in 4 months, but in- fluenced by rem. dis.	4 wks.
Result	Recovery.		
Family, & personal, history, let-out checked -	This family is rheumatic, but child herself has never had rheumatism. This is the first attack of chorea - it began in Christmas week. The Right Arm was first affected, the Right Leg followed. The Left side was then involved. Has been O.P. at St. Mary's hosp. not kept in bed. has not improved. @B. open.		
Jan. 24.	Not a very bad case - she can sit up in bed, & lay hold of an object with either hand. She can stand & walk unsupported. The right limbs are clearly more choreal than the left. Tongue protruded easily & kept out without difficulty. Hands & mouth. Muscular mems. - rather pale.		

Case note →

East London Theatre Archive



Playbill (1838), with many plays and artistes

Programme (1914), with artistes, adverts and other information



Tasks

- Transcription
- Identification of “special text”, e.g. linking from cast lists to thesaurus of actors
- Analysis of complex images
- Classification of documents (or parts thereof)
- Integration with and improvement of OCR

Advertisement
Pianos

Cast list
Julius Caesar

Cast list
Merry Wives of W.

Cast list
Hamlet

Transcribe?
Mark up?

Communities

- Different motivations
- Intrinsic interest in topic:
 - Academic / professional
 - Theatrical historians / Students
- Social focus, e.g. based on locality:
 - Local communities / Schools
 - Local history groups / Genealogists
- Interest of community rather than size

Issues to Address

- “Commensurability” of possible answers
- Definiteness of tasks
- Types of community, e.g. levels of specialisation, type of interest in archive
- Community building
- Integration with professional activities (provenance, workflow)
- Social inclusion → community benefits

Contacts

mark.hedges@kcl.ac.uk
richard.d.palmer@kcl.ac.uk