

Paul Bevan  
DAMS Implementation Manager  
Llyfrgell Genedlaethol Cymru – The National Library of Wales  
paul.bevan@llgc.org.uk

# Outline Aims

## **Primary Aim:**

Archive a representative sample of UK Political Sites related to the General Election.

## **Secondary Aims:**

- Explore collaboration between the three libraries.
- Explore the capacity for the system to archive on an 'event' basis.
- Test permissions workflows for an 'event' scenario.
- Extend the resources within archive.

# UKWAC/ National Libraries Considerations

- Overlap
  - Remit boundaries required increased definition
  - Welsh/Scottish/English/NI 'Interest'
- Permissions (Legal Context)
  - Need to seek permission for sites
  - Majority of sites unknown until late on
  - Some sites already being archived
- Software Stability
  - PANDAS
  - Previous experience cast doubt on reliability
  - Time-sensitive archiving
  - Go-live (stability)
- The role of other Partners
  - 6 Partner Consortium
  - Limited System Capacity
  - Agreement to 'hold' additional archiving during duration
- Separate Approaches

# Preparation & Planning

- Review of 'technical situation'
  - Guidelines Created
- Review of selection remit and permissions procedures (internal and in consortium).
  - 'Event' Discussion
- Review of archiving workflow, frequency and methods.

# Selection: Areas of Responsibility

- Central tenet: *on whose soil?*

It is the responsibility of the library, in whose constituency the candidate is standing, to archive the site.

- Instances of Overlap

- George Galloway

Candidate of Scottish interest standing in an English seat with substantial 'UK Political' interest. **BL remit.**

- Reg Keys

Candidate of Welsh interest standing in an English seat. Little 'UK Political' interest. **NLW remit.**

# Selection: Sample Sizes

## ■ BL

- Parties: All major and minor UK Parties
- Candidates: Representative sample of 119 candidates taken from one urban conurbation and one shire county per English & Northern Irish region.
- Other: Relevant sites related to the election (News etc).

## ■ NLS

- Parties: Initially all major and minor parties with candidates in Scotland. Later also local constituency and association websites.
- Candidates: Initially selective following a representative sample of Scottish constituencies. Later, comprehensive.
- Other: A representative sample of websites with political and social commentary on the election, including pressure groups, was also included in the selection policy.

## ■ NLW

- Parties: All national parties, all constituency-level party websites and any minor parties standing in Wales.
- Candidates: All candidates standing in Wales (having websites)
- Other: Any other germane sites, including pressure groups, media sites and 'one off' campaign sites.

# Selection: Intended Frequency

## ■ BL

- Candidates: Once
- Parties: Twice weekly
- 'Opinion Forming Sites': Once

## ■ NLS

- All Sites: Three times over election period (before, during, after).

## ■ NLW

- National Parties: Weekly
- Candidates: Twice (before, after)
- Constituency Party Sites: Once (before)
- Other sites: Once

**Note: Technical Issues**

# Permission Seeking

## ■ BL

- Initial contact by email with request to return signed form
- Secondary contact by email after 1 week giving option to 'opt out'
- Time consuming

## ■ NLS

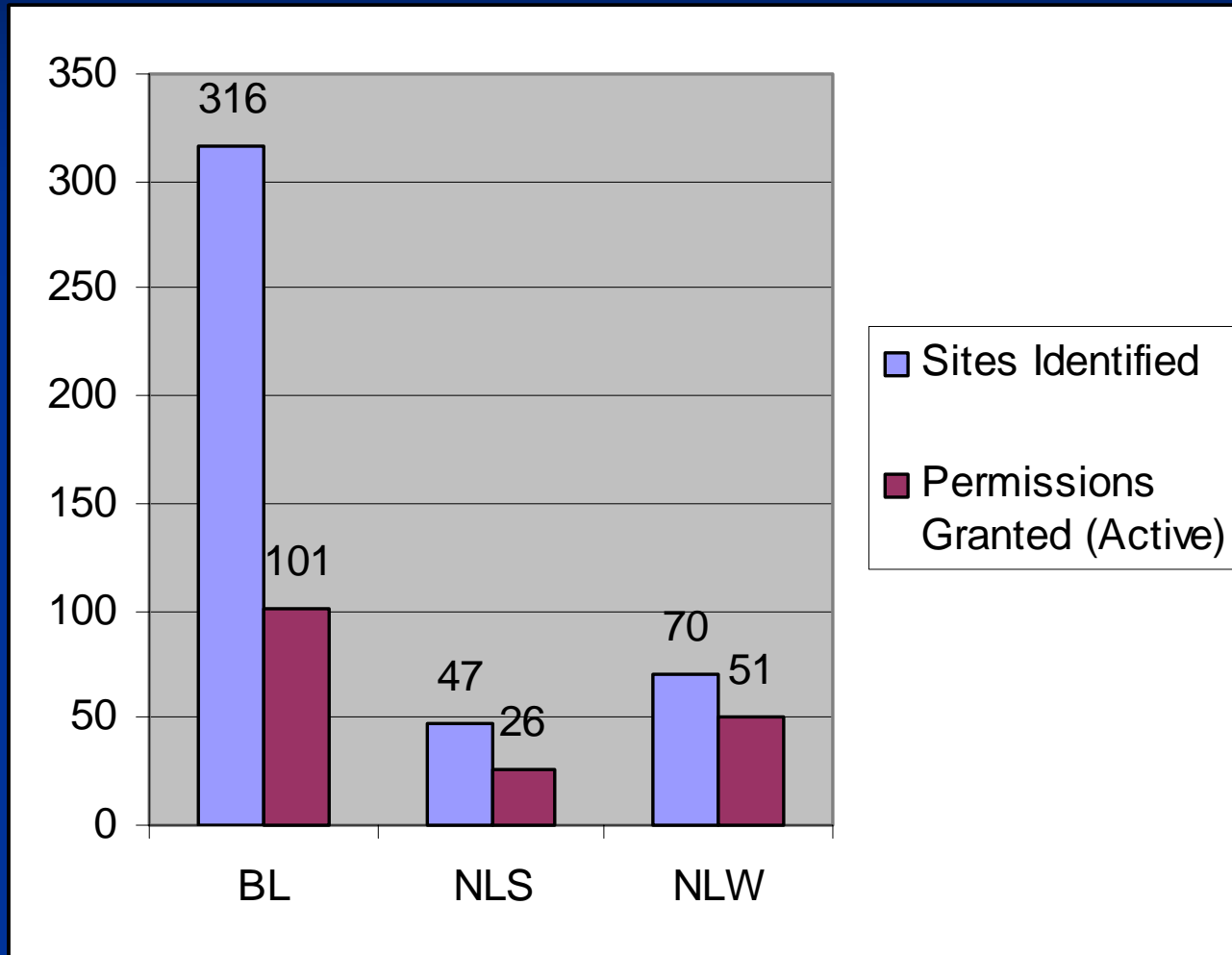
- Formal permissions sent out initially
- Email permissions accepted with formal to follow
- Time consuming but creates reusable resources.

## ■ NLW

- Some sites already sought under standard practices (eg. Main parties).
- Post or email approach for permission followed by telephone (or email) 'chase-up'. *If informal permission was given in this instance, archives were made whilst awaiting return of form.*
- Two weeks prior to election, opt-out email sent to remaining sites.
- Post-election, any sites without written permissions were approached again: No site is live without a 'permissions form'.
- Time consuming but viewed as legally essential.



# Discrepancy Between Sites Identified and Active Permissions at Time of Contact



Note: Does not include 'opt-out' or post-election permissions.

# Technical and Workflow Consideration

- 'Event' based archiving has particular requirements
- Permissions – a barrier to archiving?
- System stability & reliability paramount
  - NLS & NLW 'before, after' approach undermined.
- Website design
  - [www.conservatives.com/wales](http://www.conservatives.com/wales)
  - Already covered in UKWAC Legal Deposit libraries meeting

# Revised Expectations

- Availability of sites lower than expected:
  - NLS:
    - 32 out of 377 candidates had websites (8.5%)
  - NLW:
    - 40 constituencies with estimated 8-10 candidates + 3-4 constituency party sites per constituency.
    - Actually 62 identified (10% of estimate)
- Attempts at stability were not entirely effective
  - Site 'traps' tied up gather engines delaying queue.
  - System notification (Traffic Light) was frustrating for some staff

# Dissemination, Cataloguing and Access (ongoing)

- Majority available via UKWAC archive (as collection and individually)
- Also available via library catalogues:
  - **BL**: Single catalogue record for collection
  - **NLS**: Creation of MARC record for each site (export from Legal Deposit/Contact System)
  - **NLW**: MARC Record creation as part of future movement of PANDAS titles in to LMS

# Key Recommendations & Lessons

- General awareness of UKWAC was low
  - Pre-launch
  - Time spent explaining concept to copyright owners
- Depth should be considered along with frequency
  - Blogs are 'self archiving'
  - More time spent reviewing sites
- System feedback is essential
  - Archiving held up whilst PANDAS issues were investigated
- Permissions impact the process
  - Identification could not lead to archiving
  - Hopeful of future Legal Deposit

# Key Recommendations & Lessons

- Candidate websites are difficult to identify and come in to the race late. Some candidates were lost because of this.

So... archive as late as possible?

- The system is stable & slow *or* unstable & fast(er).
- System design (gather engines) restricts archiving
- HTTrack settings create limits (and are inflexible)
- ‘Traps’ are a major issue in a time-sensitive operation

# Further Points

- Selection  $\neq$  endorsement!
  - BL selected site contacted to ask for removal of endorsement claim
- Other interested parties:
  - LSE (especially the Liberal Democrats)
  - Oxford (sites and blogs of particular politicians)
  - Researchers seeking to understand electioneering on the Web (Wainer Lusoli, Salford)

# Report Authors

## ■ **BL**

- Jennie Grimshaw. Leading Curator, Social Sciences Collections.
- Alison Hill. Curator, Web Archiving.
- Nicola Johnson. Web Archivist.

## ■ **NLS**

- Paul Cunnea. Digital Collections Development Co-ordinator.

## ■ **NLW**

- Paul Bevan. Web Archiving Project (Systems).

Paul Bevan  
[paul.bevan@llgc.org.uk](mailto:paul.bevan@llgc.org.uk)