Cultivating community engagement through web archiving

Nicholas Worby,
Government Information & Statistics Librarian
University of Toronto Libraries
Web archiving workflow

- Selection
- Test
- Crawl
- Auto QA
- Manual QA
- Metadata Creation
Web archiving workflow

Selection

Metadata Creation

Test

Manual QA

Crawl

Auto QA

Crawl
• Labour intensive!
• Only true way to check important content
UTL’s Archive-It Government Information Collection:
• 600+ sites
• 2TB+ of data

Team:
• 2 librarians
• 4 student assistants
UTL’s Archive-It Government Information Collection:
• 600+ sites
• 2TB+ of data

Team:
• 2 librarians
• 4 student assistants

UNSUSTAINABLE!! OVEREXTENDED!!
Collaborative approaches to large digital projects ...
Search for new collaborators
Gain work experience
Develop technical skills
Engagement with professional community

Improve quality of captures
Share workload
Develop greater awareness of web archiving at UTL
iSchool QA Pilot Project

• 5 volunteers
• 1 provincial government site each
• 3 hour commitment per person
• 1 hour training session & manual
Results...

- Students took longer than expected
- Difficulty understanding scope of project
- Difficulty understanding how web archiving works
- Developed innovative strategies & made suggestions for improvements
Pilot project feedback

• More opportunities for involvement in web archiving process
• More instruction on how web archiving works
• More opportunities for feedback
• Clearer guidelines for what we are hoping to capture
Next steps

• Will run project again on upcoming mayoral election crawls
• Smaller scope
• More emphasis on training
• Involve students in other aspects of web archiving
• Explore student curated web archives
• Collaboration with iSchool faculty
Need to capture web content before migration to new site format & upcoming election

Need for a solution to preserving born digital content with a small staff

Need to capture municipal government info for researchers

Wanted to encourage municipal government use of web archiving
City of Toronto Archives collaboration

- Negotiating Permissions
- Defining responsibilities
- Training
- Support
Responsibilities

• QA work on site captures
• Metadata creation

• Regularly crawling municipal sites
• Technical support for QA process
• Metadata creation
Next steps

• Crawls of municipal site pre & post-election
• On-going collaboration on QA work & metadata creation
Conclusions

• Partners with complimentary goals
• Realistic scope
• Curb appeal
Thank you!
nicholas.worby@utoronto.ca