Archives Unleashed!
Problem: Web archives are a tremendous untapped resource for scholars, but many are unaware of how archives can be utilized.

Opportunity: Develop a forum to engage and train scholars to work with archived Web data.
Web archives are an amazing resource but access and usage are often a problem.
KEEP CALM AND HACK
It takes an army...
Format:

• 2- to 3-day event
  • **Day 1**: Introductions, networking, team formation
  • **Day 2 – 3**: Work sessions
  • **Day 3**: Final presentations & awards
Archives Unleashed 2.0
June 14-15, 2016
Library of Congress
Washington, D.C.
Putting the Plan into Action
Outcomes

Project Links

Shane Martin, Eric Oosenbrug, and Jeremy Wiebe - GraphX enhancements for Warcbase.

They have a GitHub repository here that provides insight on the D3.js side of the project

Sawood Alam and Mat Kelly - Interplanetary Wayback (ipwb)

They have a GitHub repository here that provides the initial code for integrating WARCs and the IPFS.

Nathalie Casemajor, Neha Gupta, Petra Galuscakova, Ruqin Ren, Ryan Deschamps, Rosa Iris Rodriguez Rovira. and Sylvain Rocheleau - Canadian Politics

They have a slide deck, available here in this repository.

Kim Pham, Kyle Parry, Emily Maemura, and Niel Chah - "I-know-words-and-images"

They have a GitHub repository here.

Alexander Nwala, Allison Hegel, Federico Nanni, Jonathan Armoza, Kelsey Utne, Nick Ruest, Yu Xu - "Tracking Discourse on Social Media"

They have a slide deck, available here as a Google Slides presentation. Alternatively, a PDF version is available here as well.
Day One: Socialize
Day Two - Three: Work
Day Two - Three: Work
The Final Countdown:
Talking Points by Place and Party in the last days of the 2004 Election

Research Question:
Which places and topics did the two presidential candidates mentioned most in the days leading up to the 2004 election?

Method:
1. Parse and clean the text from WARCs of party/candidate websites
2. Extracted Named Entities (candidate names, state names)
3. Topic modeling (LDA, TF-IDF)
4. Semantic Web - Google Knowledge Graph API (for next time)
Day Two - Three: Working and Sharing
How often do political parties use images of their opponents?
Forming Connections

Exchanging ideas before... and after.
Satisfaction with Outcomes

- Satisfied w Contribution
- Satisfied w Others
- Satisfied with Ideas
- Developed Useful Ideas

Dissatisfied | Neutral | Satisfied
Know your data!

What are your assumptions?
Results
- Patterns appear to validate Program on Extremism anecdotal analysis that there are "generators" and "amplifiers"
- Generated ideology scores on small sample

Next Steps
- Visualize relationship weights
- Distinguish between retweets vs. mentions
- Time-evolve relationship graphs
- Correlate with ideology scores
- Reconsider stack: Too computationally intensive for R on Mac ➔ leverage cloud instance w/ map-reduce cluster (e.g. use pig)
Archives Unleashed 3.0
February, 2017
Internet Archive
San Francisco, CA
Creating a Sustainable Model
Contact:
Matthew S. Weber
matthew.weber@rutgers.edu
@docmattweber

Archives Unleashed 2.0: archivesunleashed.com

Archives Unleashed 1.0 Projects: http://bit.ly/1ScAwaa