Archives Unleashed!

Collections as Data | September 27, 2016 | Library of Congress

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…

Library and Archives Canada

Bibliothèque et Archives Canada

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

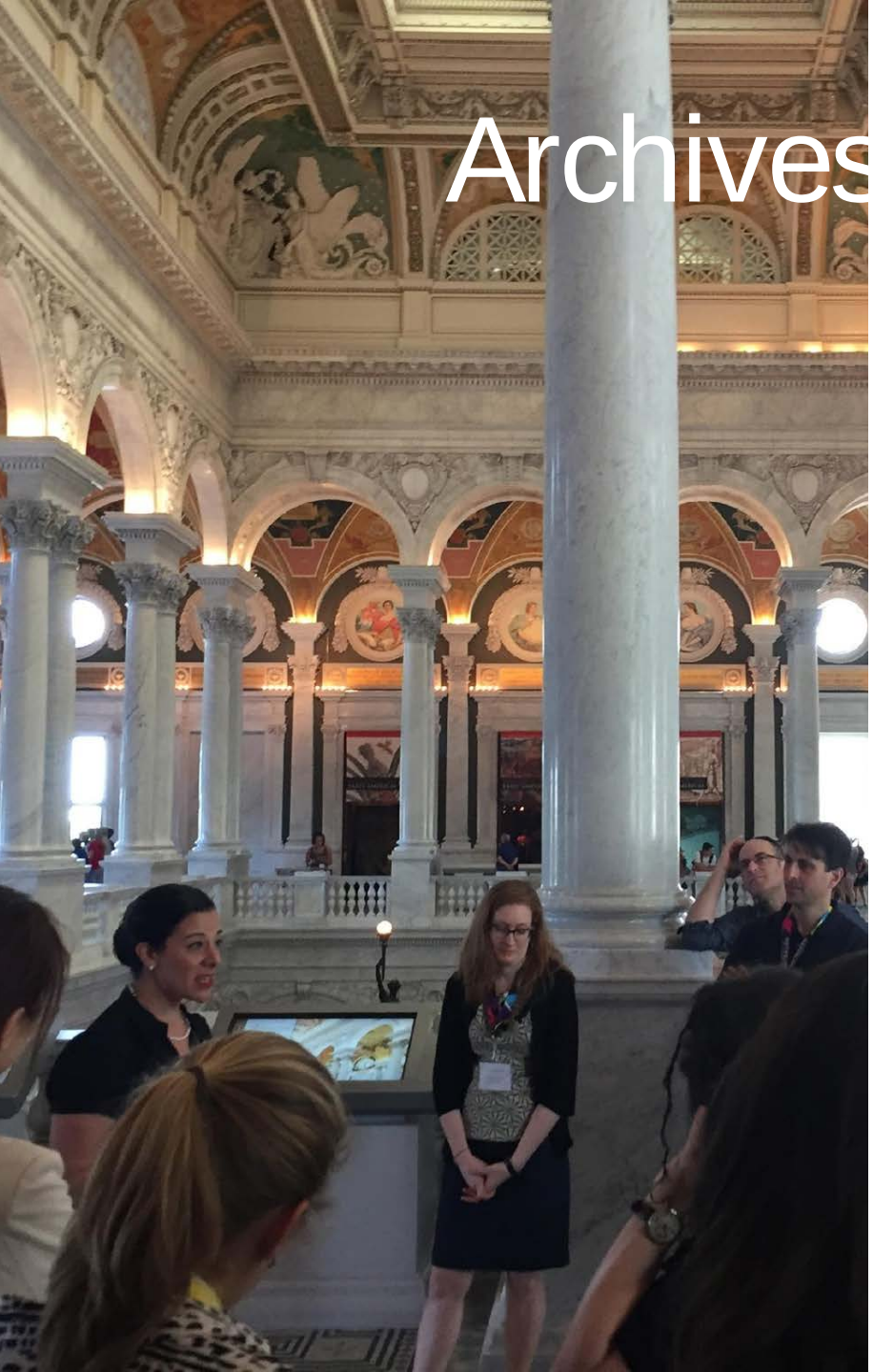
LIBRARY OF CONGRESS

Format:

- **2- to 3-day event**
  - **Day 1**: Introductions, networking, team formation
  - **Day 2 – 3**: Work sessions
  - **Day 3**: Final presentations & awards

Archives Unleashed 1.0

March 3-5, 20 16
University of Toronto
Toronto, CA

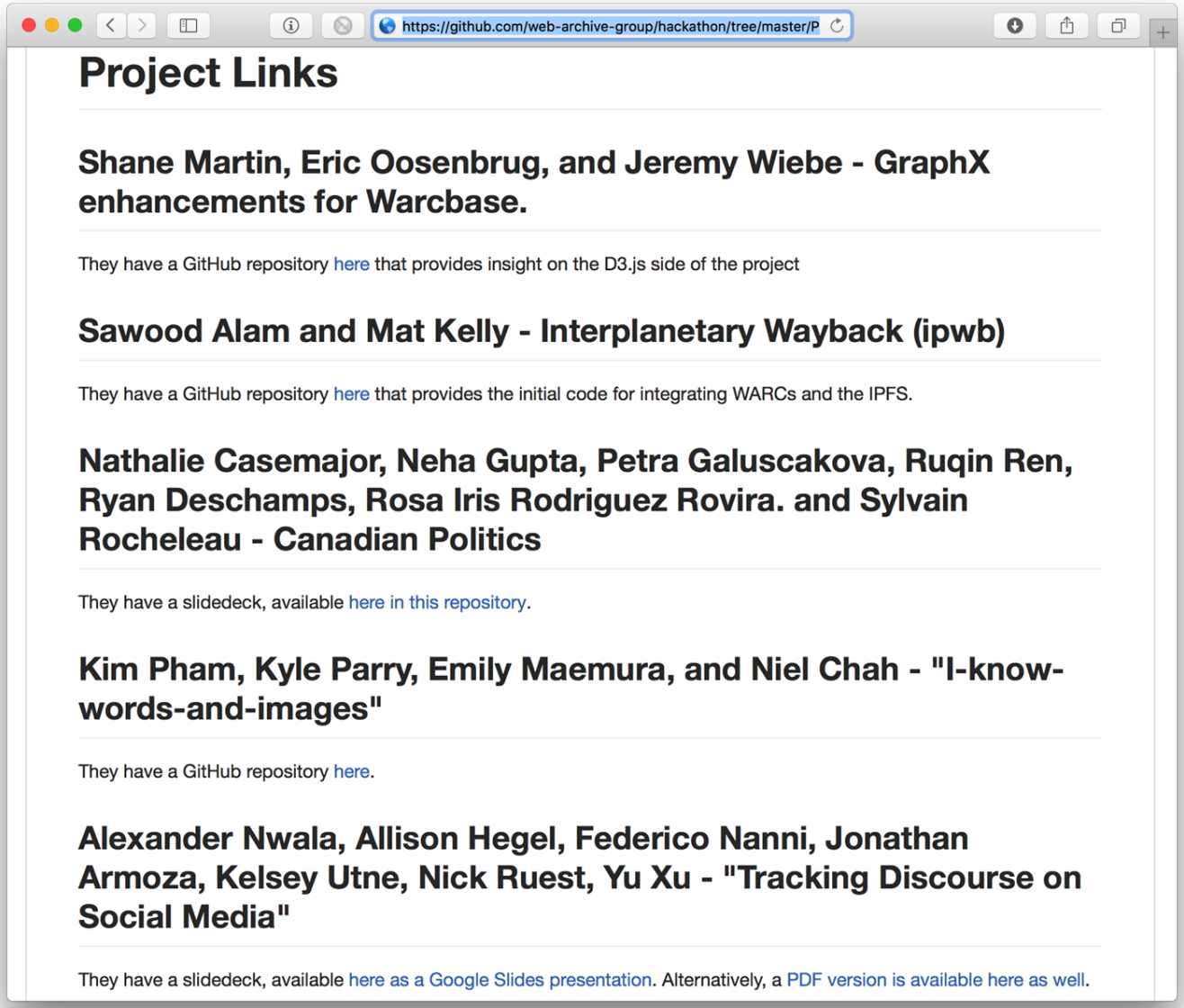


Archives Unleashed 2.0

June 14-15, 2016
Library of Congress
Washington, D.C.

Putting the Plan into Action

Outcomes

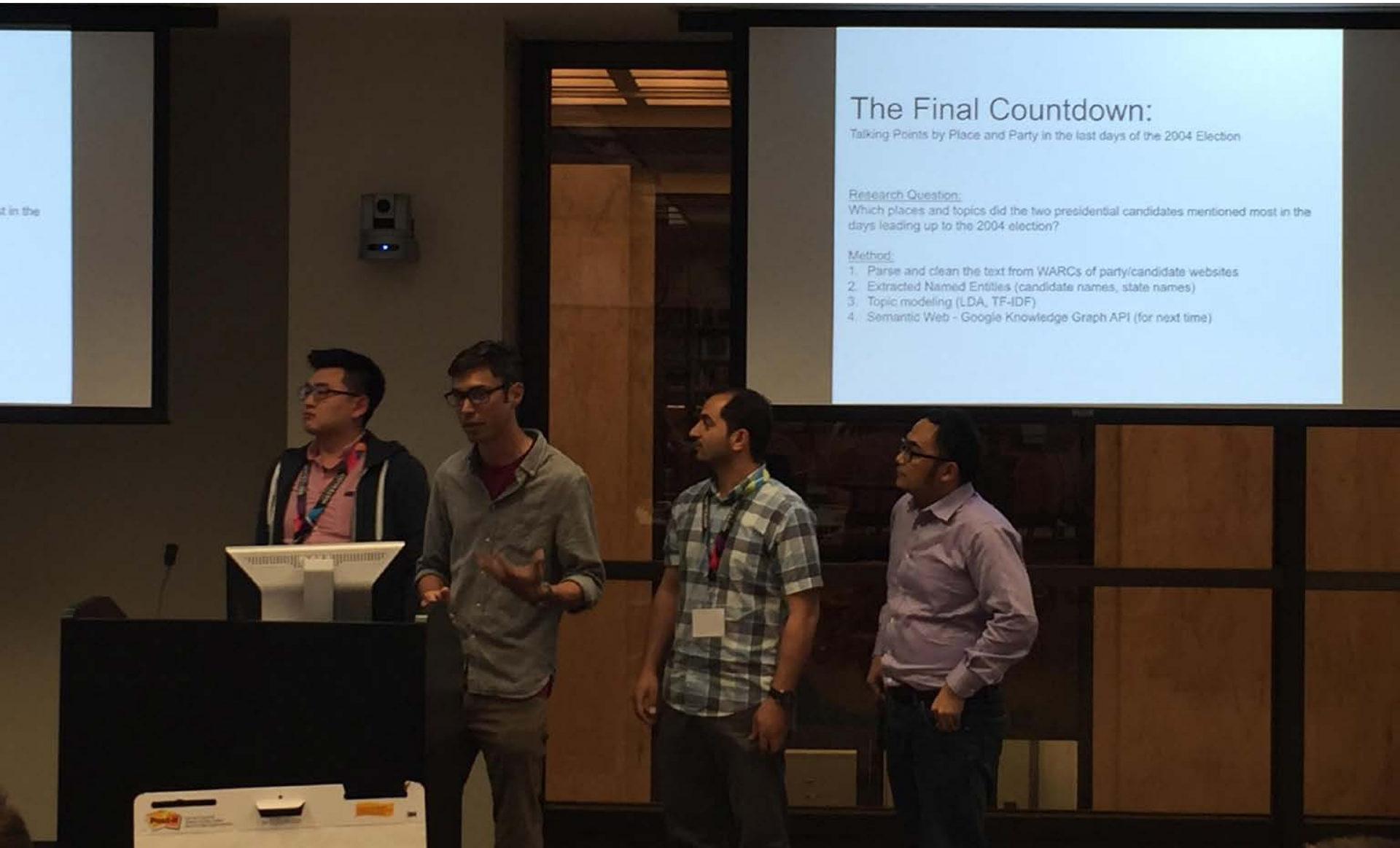

https://github.com/web-archive-group/hackathon/tree/master/P
Project Links
Shane Martin, Eric Oosenbrug, and Jeremy Wiebe - GraphX enhancements for Warcbase.
They have a GitHub repository here that provides insight on the D3.js side of the project
Sawood Alam and Mat Kelly - Interplanetary Wayback (ipwb)
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 slidedeck, 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 slidedeck, available here as a Google Slides presentation. Alternatively, a PDF version is available here as well.

Day One: Socialize

Day Two - Three: Work

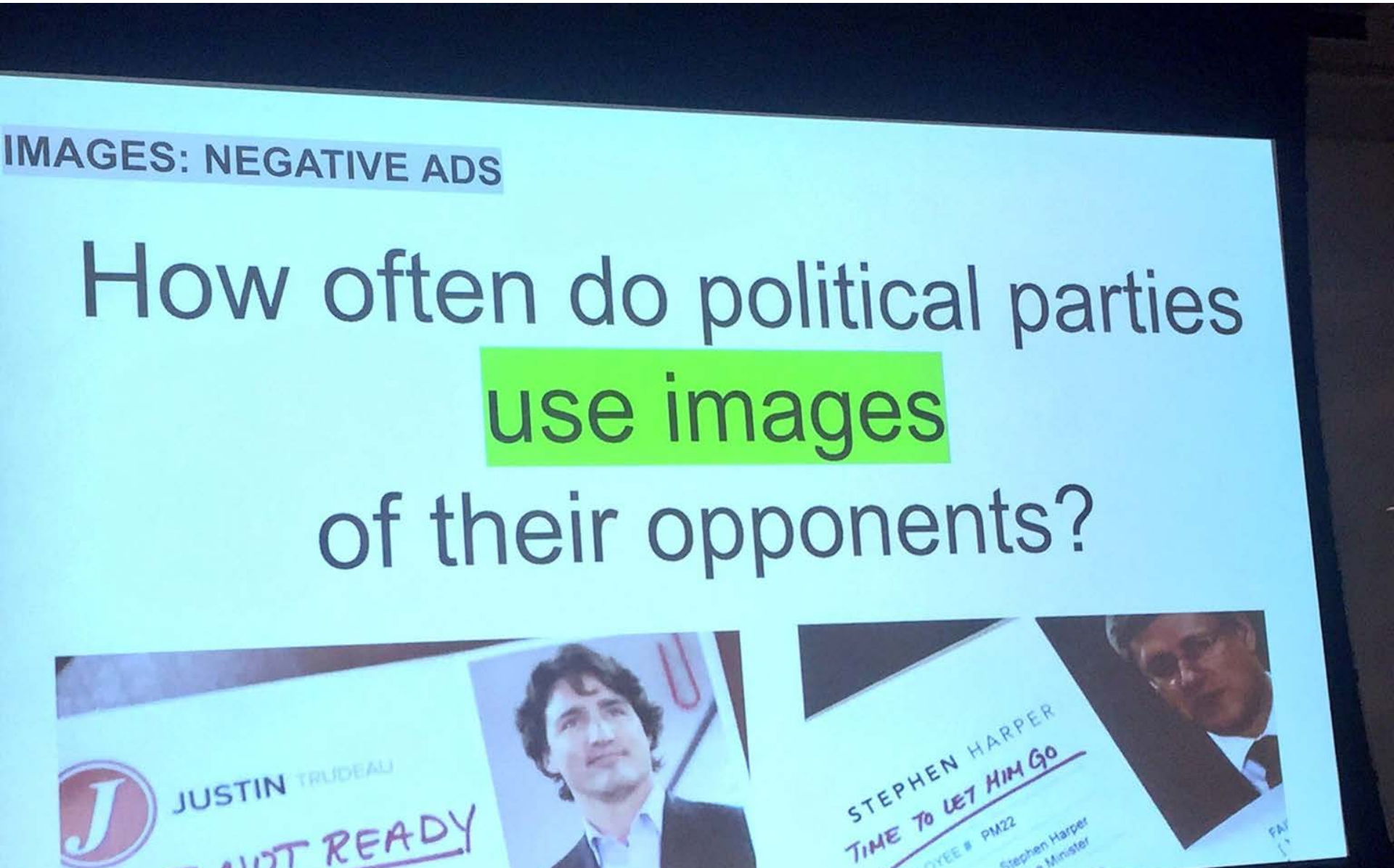
Day Two - Three: Work

Day Two - Three: Working and Sharing

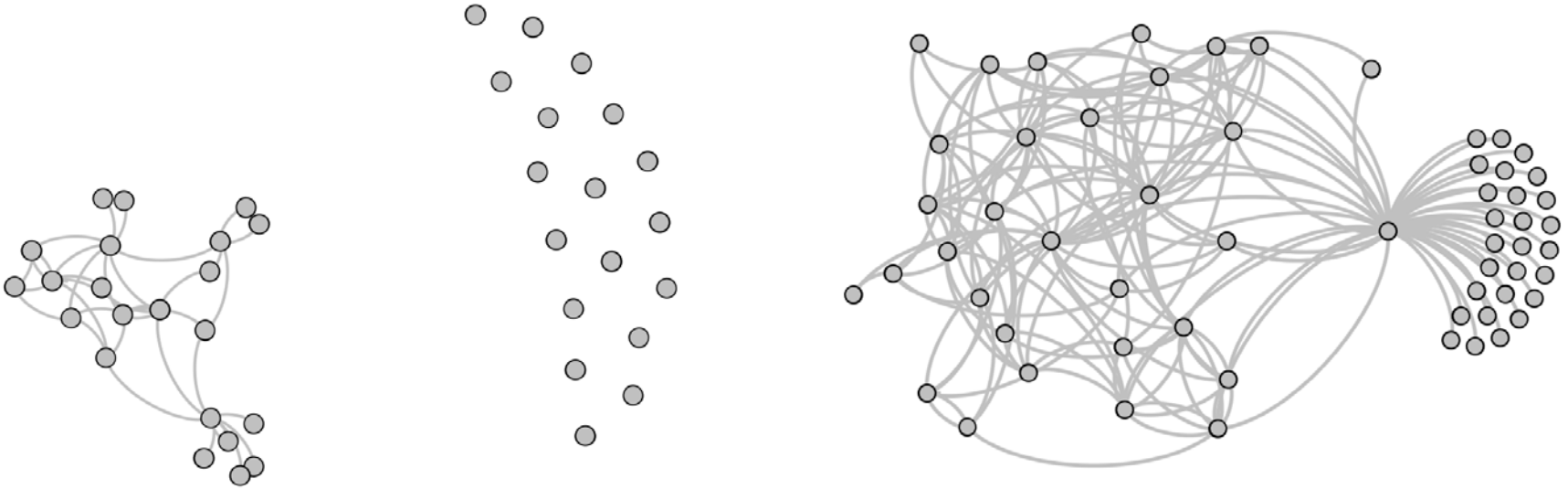


Day Two - Three: Working and Sharing

Day Three: Share!

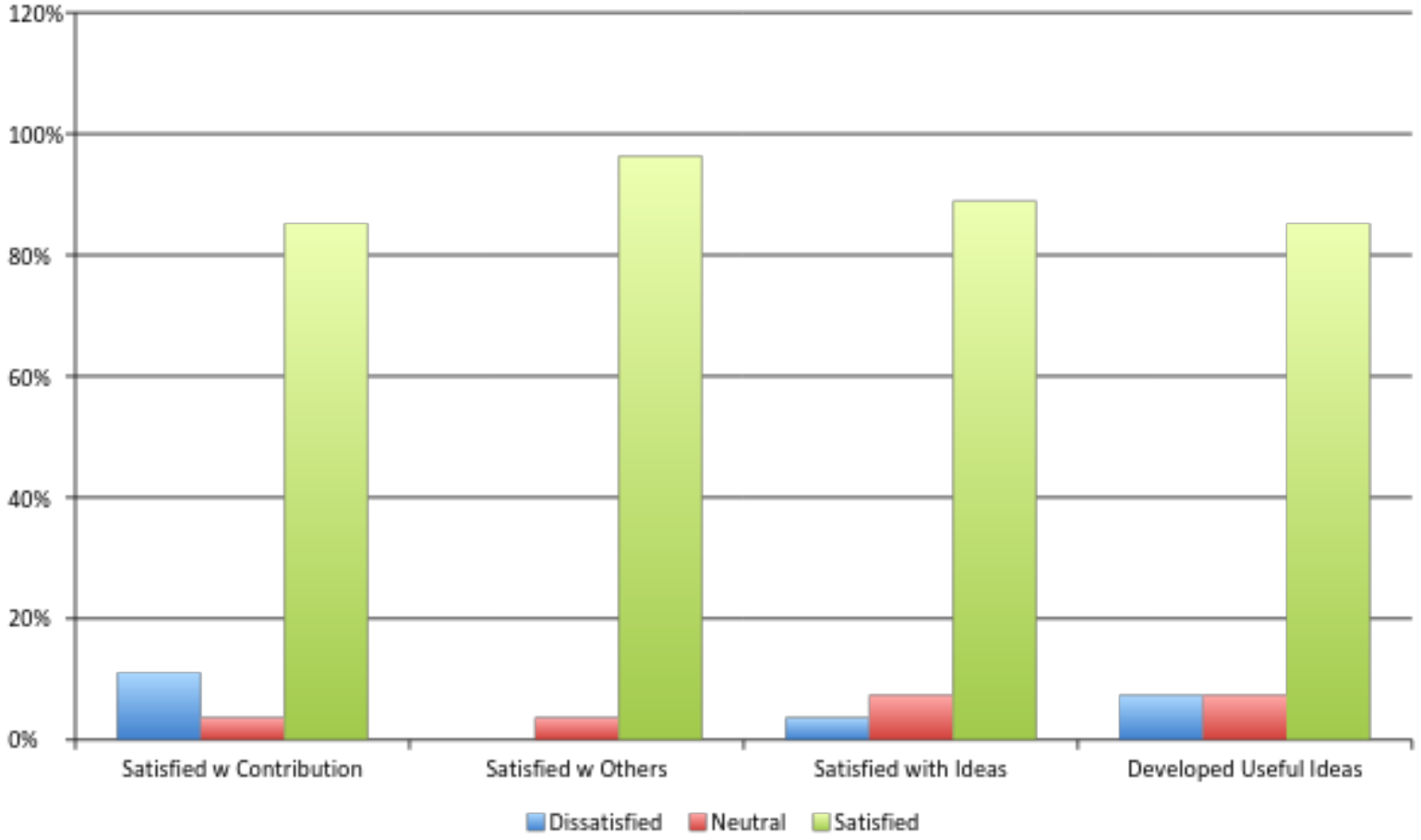


Forming Connections

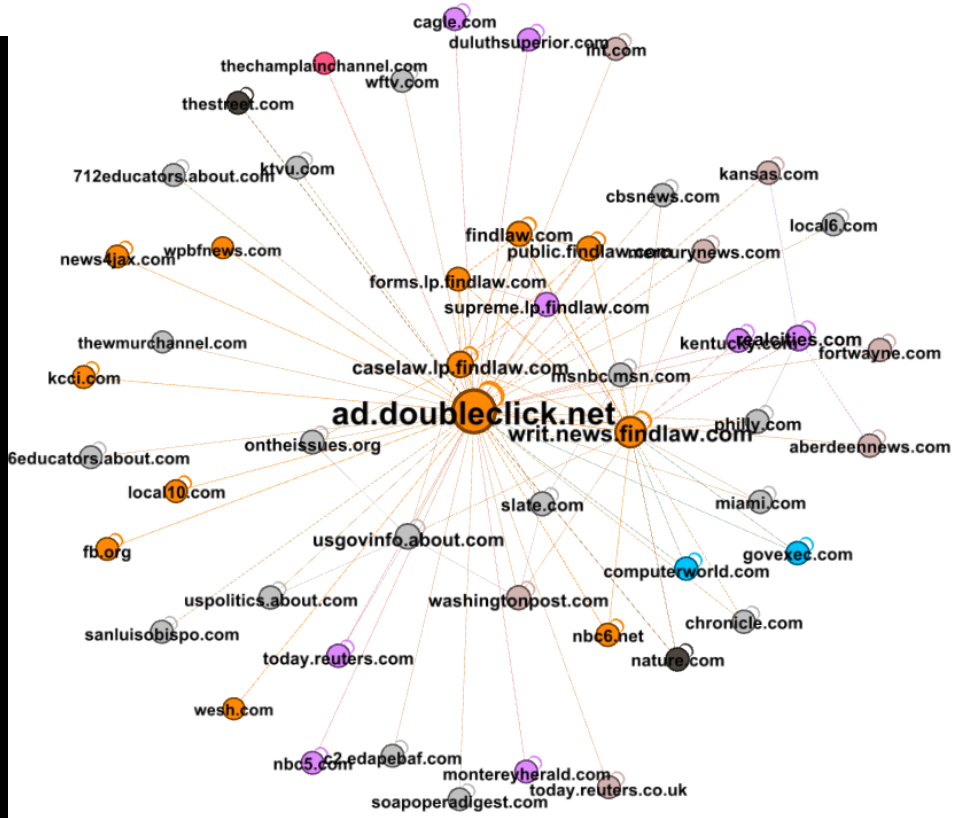


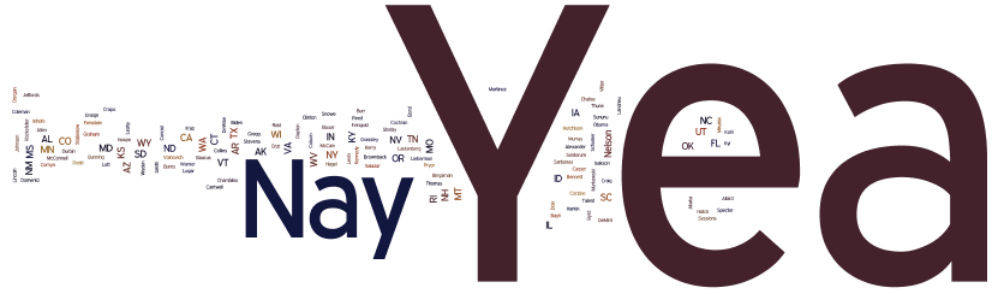
Exchanging ideas *before…* *and after.*

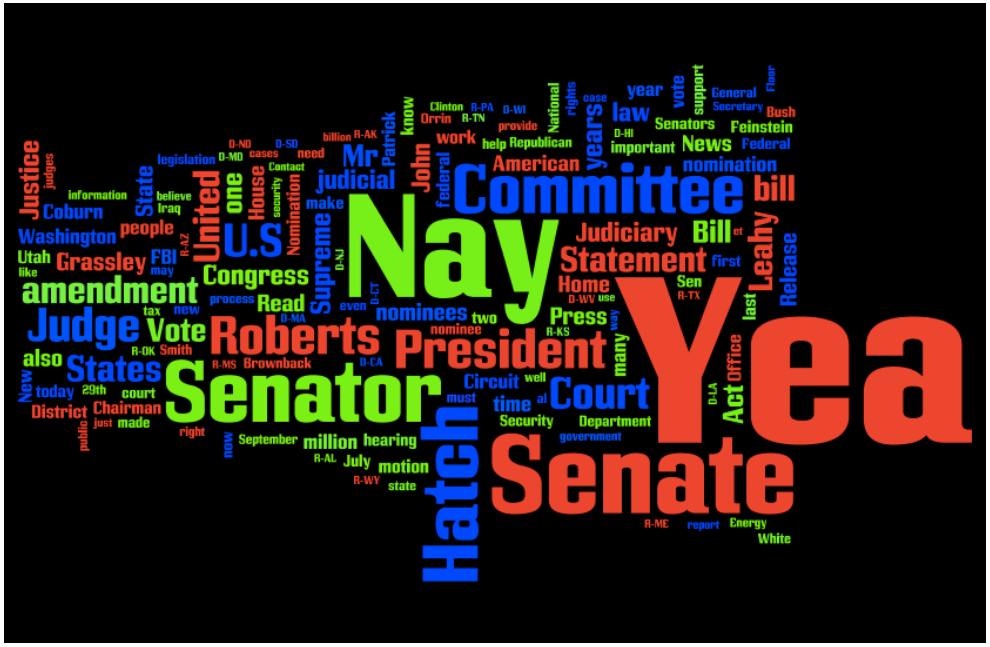
Satisfaction with Outcomes




whitehouse.gov
blogger.com
law.cornell.edu
microsoft.com
d.doubleclick.net
unionvoice.org


cagle.com
duluthsuperior.com
fnt.com
thechamplainchannel.com
wftv.com
thestreet.com
712educators.about.com
ktvu.com
kansas.com
cbsnews.com
local6.com
findlaw.com
public.findlaw.com
mercurynews.com
news4jax.com
wpbfnews.com
forms.lp.findlaw.com
supreme.lp.findlaw.com
thewmurchannel.com
kentucky.com
realcities.com
fortwayne.com
kcci.com
caselaw.lp.findlaw.com
msnbc.msn.com
ad.doubleclick.net
writ.news.findlaw.com
philly.com
aberdeennews.com
6educators.about.com
ontheissues.org
local10.com
slate.com
miami.com
fb.org
usgovinfo.about.com
govexec.com
computerworld.com
uspolitics.about.com
washingtonpost.com
chronicle.com
sanluisobispo.com
nbc6.net
today.reuters.com
nature.com
wesh.com
nbc5.com
c2.edapebaf.com
montereyherald.com
today.reuters.co.uk
soapoperadigest.com



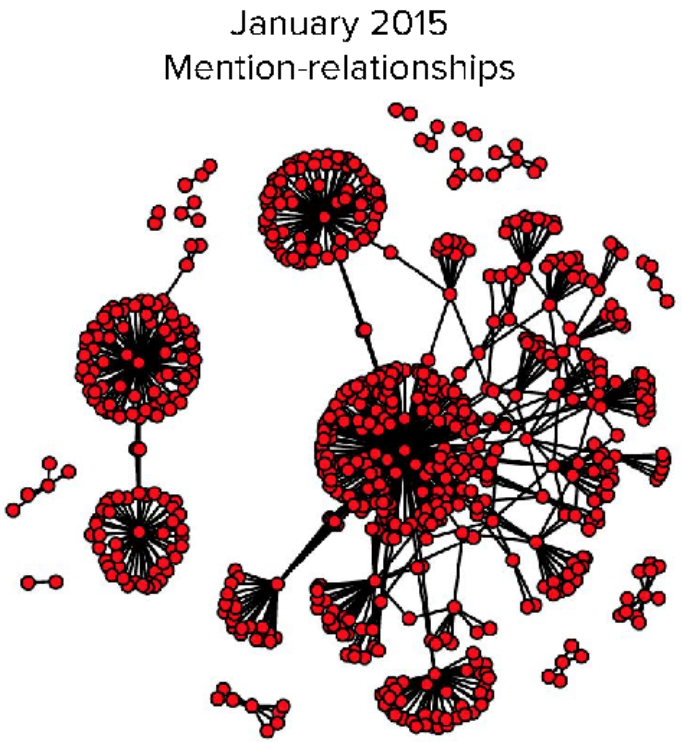


Know your data!

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

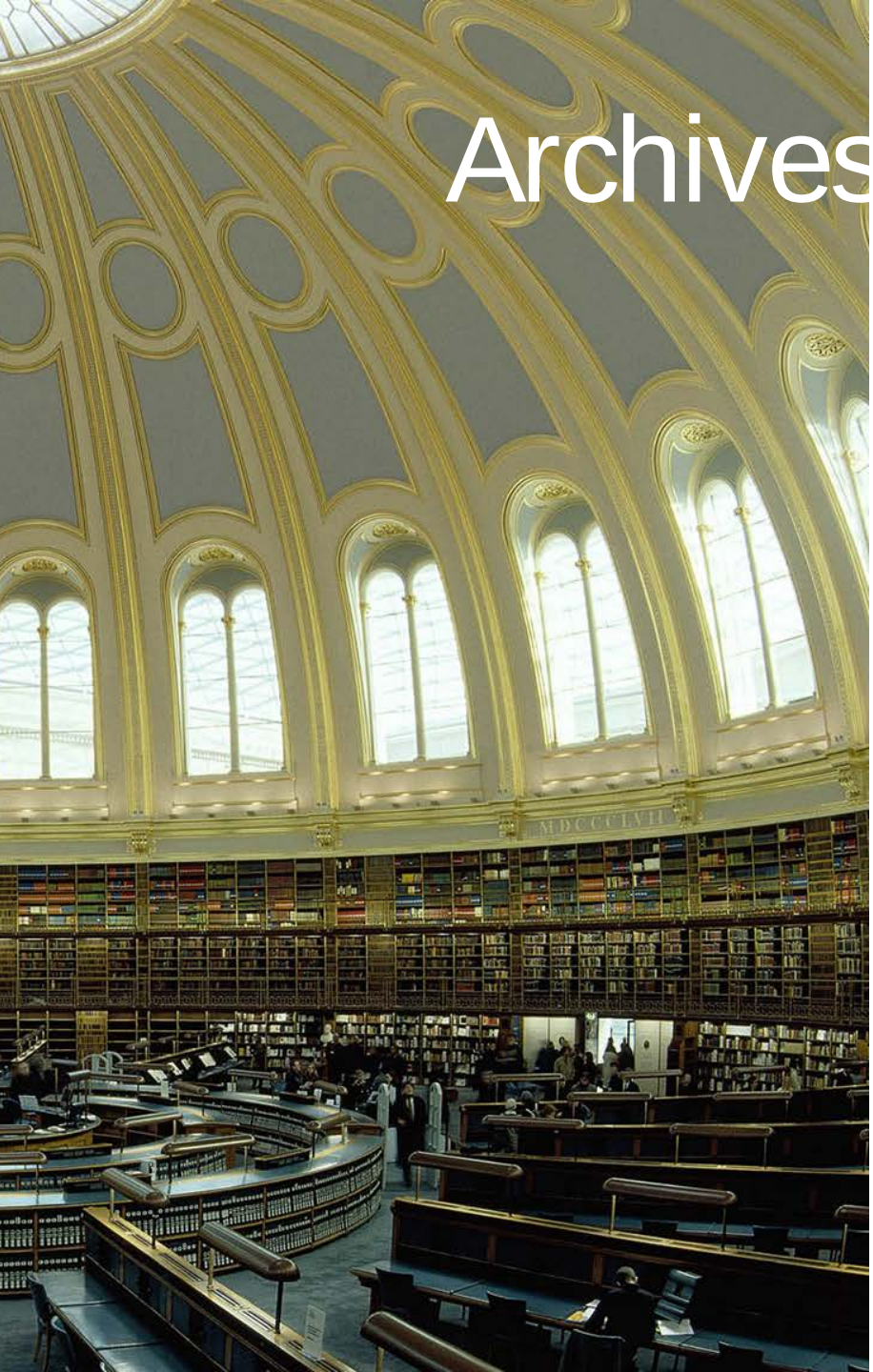


Next Steps

- 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, 20 17
Internet Archive
San Francisco, CA



Archives Unleashed 4.0

June 20 17
British Library
London, UK

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