

# eResearch needs in the social sciences

Dr. Steve McEachern  
Director, Australian Data Archive

AERO Forum  
16 March, 2016

# ADA in Brief

- The Australian Data Archive (originally the Social Science Data Archive) was set up in 1981, housed in the Research School of Social Sciences at the ANU
- Mission: to collect and preserve Australian social science data on behalf of the social science research community
- ADA preserves and disseminates over 5000 datasets from around 1500 studies, including national election studies; public opinion polls; social attitudes surveys, censuses, aggregate statistics, administrative data and many other sources.
- One collection, seven thematic archives: Social Science, Longitudinal, Qualitative, Indigenous, Historical, Crime and Justice, International
- Data holdings are sourced from academic, government and private sectors. Most Australian universities have at least one dataset deposited with ADA
- Website: <http://ada.edu.au>

# What does the social science community want to do in the next 5-10 years?

UK - Ritchie, 2009




	<- More detailed				Easier access ->		
Who	ONS Survey Mgrs	VML (ONS sites)	VML (govt. offices)	SDS	Special arrangements	Licensed data archive	Internet
Census data	Source data	Identifiable samples			Dedicated facilities	Anonymised microdata	Summary tables
Enterprise data	Source data	Identifiable data	Identifiable data	Unlinked data	Secure areas in govt. offices		Summary tables
Household data	Source data	Identifiable data	Identifiable data	Identifiable data		Anonymised microdata	Summary tables

Plus: social media (eg. online networks, sentiment analysis), administrative data (policy outcomes)  
**And that is just QUANTITATIVE social science....**

# Which eResearch approaches work well (or not) for your research methods?

eResearch approach	Usage and benefit
hosted/institutional/lab-local	Hosted: YES (e.g. Qualtrics) <b>Institutional: ???</b> Lab-local: YES (e.g. SPSS)
commercial/free/homegrown	Commercial: YES (but dilutes sharing) Free: YES Homegrown: YES
generic/discipline-specific/project-specific	Generic: YES (SPSS – Statistical Package for the <b>SOCIAL SCIENCES</b> ) Discipline-specific: YES Project-specific: LESS COMMON




# How would your community best engage with all the service providers?

- **Brokered services**, with some project-specific requirements
- Underlying eResearch layer, with tailored support for community(ies)
- Most quantitative social science can work on common infrastructure (either local or hosted)
  - Data analysis (STATA/SAS/SPSS/R), data access (ADA, ABS, Twitter, ...)
  - Increasingly data collection: e.g. Qualtrics, Limesurvey, RedCap
  - But social science data requires CURATION...
- Do not have the **capability** or the **interest** to work directly with the services:
  - What data can you provide me?
  - Where can I get a bigger laptop?
  - How can I link up X & Y?
  - Can you send me the data, and I'll analyse it on my laptop

# Data curation

- E.g. Fairfax/IPSOS poll: “Primary vote for the Australian Labor Party” or ABS “Unemployment”
- Variable:
  - Measure
  - Source question(s)
  - Response categories (male/female, male/female/other, male/female/..)
  - Method of collection (survey/admin/soc.media, online/F2F/telephone/mail)
  - Instrument (questionnaire/interview/pipeline/...)
  - Question order (what if you ask “Vote” later in the survey...)
- “Study”:
  - Investigator(s)
  - Collection process
  - Sample
- **PROVENANCE PROVENANCE PROVENANCE**
- Researchers **do not**, and more importantly **CANNOT**, provide all this information
- You cannot **link** social science data in an **automated** way **WITHOUT** this information (e.g. which ELECTORATE? which LOCAL GOVERNMENT AREA?)



# Do you have the workforce you need, and if not, where will you find it?

- **Researchers:**
  - Data analysis: probably (but lagging)
  - Data access: probably (but sensitive data is challenging!)
  - Data linkage: possibly
  - Data curation: potentially
- **Infrastructure:**
  - Compute: yes (with appropriate services e.g. security)
  - Data: potentially
  - Software development: probably (commercial providers are well-developed, less so for shared/open source/discipline services)
  - Systems: maybe (with the right eResearch infrastructure)



## Questions?

Steven McEachern

[steven.mceachern@anu.edu.au](mailto:steven.mceachern@anu.edu.au)

[ada@anu.edu.au](mailto:ada@anu.edu.au)

<http://ada.edu.au>



## So what is a data archive?

- ‘A “trusted system” that provides... an accessible and comprehensive service empowering researchers to **locate, request, retrieve and use data resources** in a **simple, seamless and cost effective way**, while at the same time **protecting the privacy, confidentiality and intellectual property rights of those involved.**’

Social Sciences and Humanities Research Council of Canada. “National Data Archive Consultation Final Report: Building Infrastructure for Access to and Preservation of Research Data in Canada” URL: [http://www.sshrc.ca/web/whatsnew/initiatives/da\\_finalreport\\_e.pdf](http://www.sshrc.ca/web/whatsnew/initiatives/da_finalreport_e.pdf) [20 November 2003].

## Future trends

- Mandated rather than recommended data archiving
- Open access to data as the default (NSF, Office of the President, data.gov(.au))
- Broader range of data types available
- Broader range of users of that data
- This is being driven by FUNDERS and JOURNALS
- “The Executive Order requires that, going forward, data generated by the government be made available in open, machine-readable formats, while appropriately safeguarding privacy, confidentiality, and security.” (Office of the President Executive Order)