eResearch and Australian social sciences

Dr. Steven McEachern
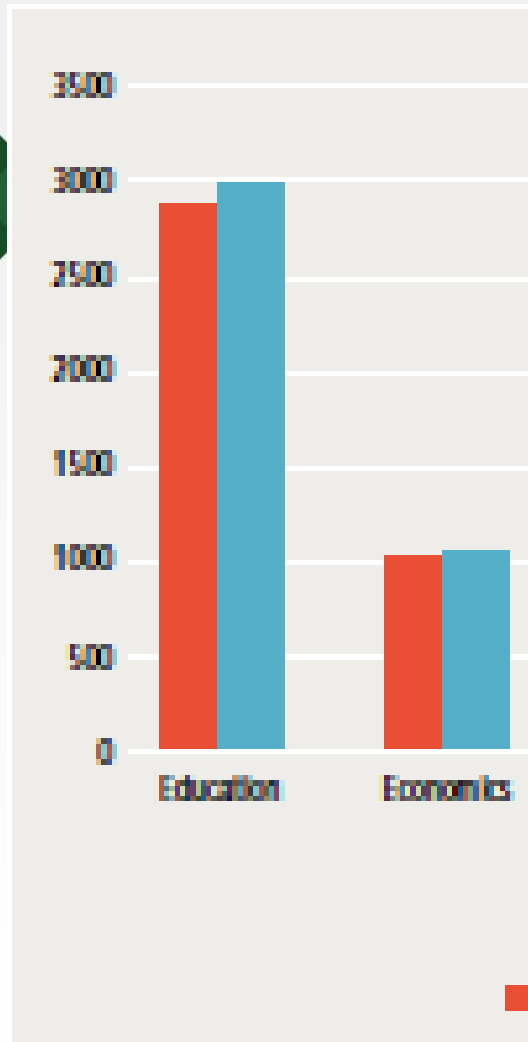
Director, Australian Data Archive

AeRO Forum, 4 May 2017

Questions on notice

- What are the research challenges you are tackling over the next 2-10 years?

- What do the eResearch needs of your discipline look like over 2-10 years? [data, compute, tools, skills, ...]

- What do you need to "thrive" in that data-rich research world of the next ten years?

- How does your community plan to participate in an open, shared, international data world?

The "HASS discipline" in Australia

Source: ARC, ERA National Reports 2010 and 2012.

The "HASS discipline" in Australia

- "HASS staff comprise approximately **52% of the total teaching workforce in higher education**. In 2012, Society and Culture staff comprise approximately 22% of total teaching FTE staff, but this represents a proportional decline from a figure of 24% in 2002.

- The ERA 2012 audit (with a census date of 31 March 2011), had **HASS research workforce numbers at 42% of the system** with Humanities and Creative Arts (HCA) at 16% and Social, Behavioural and Economic Sciences (SBE) at 26%."

- Turner, G., and Brass, K. (2014) *Mapping the Humanities, Arts and Social Sciences in Australia*. Australian Academy of the Humanities, Canberra.

- **This is not one discipline, or even two…**

What are the research challenges you are tackling over the next 2-10 years?

- There are too many areas to cover:
  - Economics, politics, sociology, criminology, education, policy studies, business management, …
- So let's look at some current examples from the journals
  - What sorts of work are people doing now

Political science

- Martinez e Coma, F. & M. Trinh (2017) How electoral integrity affects voter turnout in democracies, *Australian Journal of Political Science*, 52(1), 53-74

- http://dx.doi.org/10.1080/10361146.2016.1238869

- Does electoral integrity affect turnout? And if so, how? We analysed some variables that are closely related to electoral integrity – government actions, opposition actions, and the context in which the election is held – and find significant impact on turnout. We argue that higher turnout is often found in elections with higher electoral integrity. We tested our claims using data for over **700 elections covering 85 democracies for the 1950–2008 period**. Results reveal that both boycott and election-related violence decrease turnout, but the effect of the former is substantially higher. We also find that, contrary to initial expectations, governments' harassment of the opposition and the occasional banning of parties actually increases turnout.

Sociology

- With geopolitical concerns surrounding the rise of militant, transnational groups who draw on Islamic texts for legitimacy, the place of Islam in western societies has become a source of anxiety, fear and suspicion. The central concern is whether Muslims living in the West have the capacity to become fully active citizens. This article uses **quantitative and qualitative methods** to examine whether Islamic religiosity is a predictor for civic engagement and active citizenship among Muslims living in Melbourne, Australia. The findings show that organized religiosity can be a strong predictor of civic engagement, countering the discourses that demonize Islam as a source of radicalization and social disengagement. While the findings show that suspicion of divisive forces and lack of trust in public institutions might prevent some young Muslims from engaging in formal political participation, grassroots civic engagement enables Muslims to demonstrate care and feel like active citizens of the Australian community without compromising core religious values.

- "The qualitative data was collected via **semi-structured interviews of approximately one hour** in length, conducted with 48 participants (Female = 22; Male = 26) recruited through snowballing methods. Four focus groups were also conducted with religious and community groups. The data set that we used for the quantitative analyses draws on **an original survey of practising Muslims in Melbourne** (N = 96)."

Economics

- Jeff Borland and Michael Coelli (2016) Labour Market Inequality in Australia, *Economic Record*, 92(299), 517-547. DOI: 10.1111/1475-4932.12285

- "This article reviews developments in labour market inequality in Australia. First, descriptive information on changes in labour market inequality and on the causes of those changes is integrated with a summary of findings from recent research. Second, the effect of changes in labour market inequality on income inequality is evaluated. Third, evidence on differences in earnings and employment outcomes between specific groups – by gender, Indigenous status and country of birth – is considered. Finally, some suggestions for future research are presented."

- "**A variety of data sources is available** to describe labour market earnings inequality in Australia …"
  - Income Distribution Surveys (IDS)/Surveys of Income and Housing (SIH) – ABS
  - Labour Force Surveys (LFS) – ABS
  - Employer Surveys (ES) – ABS
  - The HILDA Survey – Department of Social Services/Melbourne Institute
  - Australian Tax Office (ATO)

Education

- Jenny Chesters & Anne Daly (2017) Do peer effects mediate the association between family socio-economic status and educational achievement? *Australian Journal of Social Issues*, 52(1), 63-77. DOI: 10.1002/ajs4.3

- Differences in levels of academic achievement according to socio-economic status (SES), and parental education in particular, have been a persistent feature of Australian education systems. Young people with highly educated parents are more likely than their peers with low-educated parents to attain high levels of achievement at school. Students with low levels of achievement are less likely than their high achieving peers to complete Year 12 and are more likely to experience negative post-school outcomes. The SES of the neighbourhood, and in particular, the school attended, has also been found to have an effect on levels of both academic achievement and attainment. For this paper, we conduct analyses of **National Assessment Program – Literacy and Numeracy test scores** for four cohorts of secondary school students attending government schools in the Australian Capital Territory to examine the associations between parental education, school attended and levels of educational achievement. Our findings show that students with university-educated parents achieve at much higher levels than their peers with low-educated parents and that attending a school with a higher proportion of students from educationally disadvantaged families has a negative effect on educational achievement.

Themes

Foreign policy
- Democracy
- Civic engagement
- Violence
- Terrorism
- …

Economic policy
- Labour markets
- Economic policy
- …

Social policy
- Inequality
- Social class / Socio-economic status
- Parenting
- Child development
- Racism
- …

Future trends

- Mandated rather than recommended data access
- 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)

What do the eResearch needs of your discipline look like over 2-10 years? [data, compute, tools, skills, ...]

- **Data** and **Skills** are key challenges:
  - How do I **access** data – particularly about individuals at a "unit record" level
  - E.g. government administrative records, health records, biomarkers
  - How do I **analyse** the data – increasingly complex models
  - What happens when I get to "big data": administrative records, social media data
- **Tools** are well-established in most areas
- **"Big"** Compute is not generally an issue
  - But note increasing "big data" demands
- **"Secure"** Compute and Storage often is an issue

What do you need to "thrive" in that data-rich research world of the next ten years?

ADA AUSTRALIAN DATA ARCHIVE

Short term:
- Access to data
  - How do you access NAPLAN
  - How do you access the Twitter firehose
- Linking data
  - NAPLAN records to parent or child information
  - Immigration information to census information
- The skills to analyse that data
  - Linked and longitudinal data
  - Text, audio and video data
  - What can you do with that data once you have it?
- Better support for the research lifecycle:
  - Conception -> Funding -> Ethics -> Data collection -> Analysis -> Reporting
- Better means for documenting and reporting research
  - Fraud, Reproducibility
  - Publications -> Data -> Analysis code -> Software -> …???

Supporting the research lifecycle

Australian Government
National Health and Medical Research Council

Australian Government
Australian Research Council

Conception ->
Funding ->
Ethics ->
Data collection ->

Welcome to the Human Research Ethics Application (HREA)

Sign

User

Pass

HREA Help

The **Dataverse** Project

Reuse -> (Recon

Don't want an account? *Guest access*

THE POWER TO KNOW®

SPSS

What do you need to "thrive" in that data-rich research world of the next ten years?

Long-term:

- Generating or accessing new forms of data
  - "Passive" observation (e.g. accelerometers in iPhones)
  - Emerging social media
- Accessing old forms of data
  - Data extraction from historical records (e.g. convict records from Tasmania)
  - 100 years of election results
  - 200 years of census results (1000 years in some countries)
- Developing appropriate means for analysing these new forms
  - Social network models
  - Semantics of content in audio and video

Founders and Survivors

ADA AUSTRALIAN DATA ARCHIVE

- "Founders & Survivors is a multi-university and public collaborative project that is building a transnational and intergenerational dataset of life courses generated from the UNESCO recognised convict records of Tasmania.
- Mass digitization and archiving online of over 100,000 images
- Manual scholarly transcription of hand-written content from those images
- Building of a prosopography database (think a linked data system for 6000 individuals over 100 years). This comprises a relational genealogy database integrated with an XML (BaseX) source database.
- Individual life histories are compiled dynamically from diverse sources, linked by a combination of machine matching and human judgment, and managed by an independent link management module.
- Using Google Docs over 50 online volunteers crowdsourced the convict genealogies and coded the data.
- Manual linkage and scholarly verification remained essential for the collation of prosopographical data and manual coding was necessary for statistical analysis."
- (http://dx.doi.org/10.1007/978-3-319-19884-2_14)
- http://foundersandsurvivors.org
- Longer term aims:
  - Link to World War One Australia Imperial Force Service Records
  - Study the multi-generational effects of nutrition and transportation on descendant health outcomes

Layout: Trial & Transportation L... ▼ | View As: ☐ ☰ ▦ | Preview | Aa | Edit Layout

Welcome to the **Female Convicts in Van Diemen's Land** database

Convict ID	497	Police Number 1000
Convict Given Names	Anne	
Convict Family Name	Corry	
Native Place		Clare
Country	Ireland	

Ship to VDL Australasia | Voyage

Departure Date 26 June 1849
Departure Port Dublin, Ireland Ireland
Arrival Date 29 September 1849
Arrival Place Hobart VDL

HELP! How to use the database.
Show All Records
Find
Next
Previous
Contributors
Sources Help!

Aliases:

Given Names	Family Name	Order	Source
Ann	Corry		TAHO index
Ann	Curry		CON19
Anne	Gallagher		married name

Ship/s to Other Colonies | Voyage
Departure Date
Departure Port
Arrival Date
Arrival Place

Trial & Transportation

Trial & | Convict | Locations/ | Colonial | Probation | Levels of | Surgeon's | Bank | Police Gazette | | Absconder | | Punishment | Sources

---

Warnock Isabella

Tried Antrim 8th April /45 10 Yrs

Embarked

Arrived 3rd December, 1845.

Transported for Larceny from the Person. Gaol report Larceny. Stealing a Watch. Single. Stated this offence Stealing a Watch prosecutor Mr Campbell. Once for a cap 1 month and for a watch 4 months and acquitted for Clothes. Single. Surgeons Report

Trade.	Height.	Age.	Complex.	Head.	Hair.	Whiskers.	Visage.	Forehead.	Eyebrows.	Eyes.	Nose.	Mouth.	Chin.	Native Place.
Dress maker	5'3½	24	Sallow	Long	Dark Brown	—	Long	Low	Brown	Grey	Cocky	Wide	Broad	Dublin

Marks

Arrived Free?	no	Number Of Times Sentenced to Transportation 1
Multiple Records?	no	Previous Sentence Of Transportation no

Grangegorman Depot

Information on this page is primarily from the conduct record.

75 ☐ ⊕ ☐ Browse

How does your community plan to participate in an open, shared, international data world?

We have done for a long time:

- 1960s: establishment of social science data archives in the USA, UK, Germany, Norway (1971), and progressively across OECD countries
  - Australia: proposed in late 1970s, **ADA established in 1981**
- 1970s: establishment of international collaborative organisations
  - CESSDA: Council of European Social Science Data Archives
  - IFDO: International Federation of Data Organisations in the Social Sciences

General aims:

- **Sharing of research data** across the social sciences
- Development of communities, standards and data sharing agreements **within and between countries**
- Support for **international research programs**

Why? Many problems we are interested in are **international in scope**

International research programs

- 1948: First "American Election Survey"
- 1972: First "General Social Survey" (U.S.) and "ALLBUS" (Germany)
- 1983: British and Australian Social Attitudes Surveys
- ISSP: International Social Survey Program
  - 57 countries annually (min. 1000 surveyed per country)
- ESS: European Social Survey
  - 36 countries every 2 years (~2000 surveyed per country)
- WVS: World Values Survey
  - ~100 countries every five years (min. 1000 surveyed per country)
- SHARE: Survey of Health and Retirement in Europe
  - 123,000 people, 27 countries, every 2-3 years

ADA AUSTRALIAN DATA ARCHIVE

Where are the social sciences infrastructures heading internationally?

Europe: now have three social science ERICs

- European Social Survey
- Consortium of European Social Science Data Archives
- Survey of Health and Retirement in Europe
- SERISS: integrated program to coordinate shared needs across the three infrastructures

USA:

- Data as infrastructure: GSS, ANES

Australia:

- ???

National and International organisations

ADA AUSTRALIAN DATA ARCHIVE

Data collection and research is both funded by and produced by:

Australian:
- Australian Bureau of Statistics
- AIHW, AIC, AIFS
- ATO
- Departments of:
  - Social Services
  - Education
  - Human Services
  - Health
  - ...
- **Note Productivity Commission report**

International
- World Bank, Asian Development Bank
- OECD
- United Nations (UNESCO, UNPF, WHO)
- International Monetary Fund
- USAID, AusAID, NZAID, ...
- Etc.

Questions?

Steven McEachern
steven.mceachern@anu.edu.au
ada@anu.edu.au
http://ada.edu.au