ADA-VL: A virtual laboratory for Australian social science

Dr. Steve McEachern
Deputy Director, ADA

School of Sociology seminar series
October 2011
Presentation Overview

1. ADA - the current state of play
2. Recent developments
3. Where should we go next?
4. An ADA-Laboratory
5. Future directions
1. ADA – the current state of play
ADA in Brief

• The Social Science Data Archive (now ADA) was set up in 1981, housed in the Research School of Social Sciences, Australian National University, with a mission to collect and preserve Australian social science data on behalf of the social science research community.

• Now includes nodes at University of Melbourne, University of Queensland, University of Western Australia, University of Technology Sydney, with infrastructure provided by the ANU Supercomputer Facility.

• The Archive holds some 2400 data sets, including national election studies; public opinion polls; social attitudes surveys. Data holdings are sourced from academic, government and private sectors.
The original research community needs identified by the ASSDA Advisory Panel to be addressed by the ASeSS project were as follows:

1. A coherent single point of access for nationally significant social science and associated humanities resources, including access for researchers, students, government bodies, and other external agencies.

2. Reliable access to the major national social surveys.

3. Management of a diverse range of data forms needed to help answer research questions across these different forms: eg: unit record data, qualitative data, economics data, including a high level of data documentation that allows researchers to quickly identify its relevance and quality for research purposes.

4. Easy access to specialised collections, eg: topic based data, such as data relating to ageing; colonial data; indigenous data.

5. Provide fast search across all this data.

6. Easy access to data analysis tools, including the development of advanced analytical and visualisation tools and capability (outside of commercially available products) that provide additional value to the data archives and support the ‘unlocking’ of otherwise inaccessible data sets of national significance.

7. Computational modelling, expertise and resources including computationally expensive statistical packages.
Current Approach

• Core archive website:
  – http://www.ada.edu.au

• Sub-archives focussed on specialised thematic or methodological areas

• “Add-on” systems for complex analysis or visualisation tasks:
  – Nesstar
  – GIS: http://gis-test.ada.edu.au
  – Longitudinal visualisation: Panemalia
  – Historical census data: http://hccda.ada.edu.au
2. Recent developments
Technical developments

- Cloud computing infrastructures (Amazon, NASA Openstack, …)
- New analysis environments
  - Rstudio
  - Dataverse
- Data collection systems
  - Online services
  - Open source services
- Integration
  - DDI
  - SDMX
Methodological developments

• Geospatial analysis
• Longitudinal analysis
• (Multilevel analysis)
• Computational qualitative methods
• “Big data” analysis
Policy and funding developments

- National Collaborative Research Infrastructure Strategy
  - Dept of Innovation, Industry, Science and Research (and Tertiary Education from 2012)

- NCRIS Platforms for Collaboration/National e-Research Taskforce:
  - Funded ADA developments 2008-11

- NeCTAR: National eResearch Collaboration Tools and Resources
  - National Server Program
  - *National Research Cloud (based on OpenStack)*
  - *Virtual Laboratories program*
  - eResearch Tools
3. Where should we go next?
Where to from here?

- Audio-visual (LIEF 2011-12)
- NeCTAR program: Data integration
  - Secure data access (administrative data, data linkage)
  - Qualitative data documentation and analysis
  - Historical/time series spatial analysis
  - Geospatial and temporal data integration
  - Integration across content types – eg.
    - Election results, poll results, candidate surveys
    - Census, survey and administrative data on a topic (eg. crime)

- Program includes focus on eResearch tools, and virtual laboratories
Interest in new ASSDA services (n=99-101)

- Census data: 45.00%
- Longitudinal data (eg. panel studies and time series): 41.40%
- Data related to elections and politics: 36.60%
- Data related to public opinion polls: 33.00%
- Links to international databases (eg. OECD, United Nations, World Bank): 27.70%
- Secure services for accessing confidential survey and administrative data: 24.00%
- Linkage of multiple data sources: 21.80%
- Qualitative data (eg. interviews, focus groups, images, documents): 20.20%

0 - Not at all interested  1  2  3  4  5 - Very interested
4. An ADA virtual laboratory
What do we need?

• More streamlined access to data
• Data integration
  – “Data mashup”
  – Data linkage
• Access to complex analysis capabilities:
  – Desktop
  – High performance computing
  – Remote access
• Secure analysis environment
  – Remotely accessible
Use cases

• Census data access
• Survey data collection and analysis
• Geographical data integration
• Longitudinal and time series dataset creation
• Data export of higher-risk content
A VL model
5. Future directions
Where to from here?

- Funding application: June 2012
- Start-up: September 2012
- Project completion: December 2013
Questions or comments?

For further information

Web: http://www.ada.edu.au

Email:
steven.mceachern@anu.edu.au