



# Data Archiving in Australasia: Lessons from across the ditch

Sophie Holloway  
Australian Social Science Data Archive

**BRCSS Virtual Seminar Series**

**16 August 2007**



# Presentation Overview

- The Digital Preservation Imperative
- E-Research and Infrastructure
- Who pays for preservation/sustainability
- ASSDA's Structure

# ASSDA at Establishment



- ASSDA was set up in 1981, housed in the Research School of Social Sciences at the Australian National University
- It was originally set up as a part of ACSPRI - Australian Consortium for Social and Political Research Incorporated to bring overseas data into Australia and to preserve Australian data.
- ACSPRI is Australia's peak body for social research methods and has been running statistical training courses since the 1980's and has recently started running Research Methods Conferences (see [www.acspri.org.au](http://www.acspri.org.au))

# ASSDA's Collection



- The Archive holds some 1500 studies, most notable holdings are national election studies; public opinion polls; social attitudes surveys.
- Data holdings are sourced from academic, government and private sectors.
- The Archive also acts as a custodian for data from New Zealand, Indonesia and other Asia/Pacific nations.

# Cultural Changes to the Operating Environment



For most of its history, ASSDA operated in a highly centralised fashion serving a small, fairly tight knit community of users - who were also willing depositors.

In the 1990s there was a cultural shifts in attitudes towards data

- ABS data commercialised, data now see as a potential source of profit
- Academic community began hoarding research collections
- Deposits started dropping

# Changes to Data Collections



- In the 1990s social science data available in electronic form expanded and became increasingly diverse. We saw the rise of:
  - Qualitative data
  - Administrative data

And learned that different data types needed a different approach

- We also saw the rise of totally different research
  - Web pages
  - Visual images

# The Concept of Digital Heritage



- The start of the new century saw the recognition that most research tools and records of culture were now being born digital
- While digital information is easy to record, copy and store, it is also easy to lose and it can become obsolete very quickly
- Increasingly, the international community became aware of this problem and through UNESCO issues a Charter on the Preservation of Digital Heritage where member states were obliged to preserve research data

# UNESCO Charter on the Preservation of the Digital Heritage, 2003



## Article 10 – Roles and responsibilities

Member States may wish to designate one or more agencies to take coordinating responsibility for the preservation of the digital heritage, and to make available necessary resources. The sharing of tasks and responsibilities may be based on existing roles and expertise.

Measures should be taken to:

- (a) urge hardware and software developers, creators, publishers, producers and distributors of digital materials as well as other private sector partners to cooperate with national libraries, archives, museums and other public heritage organizations in preserving the digital heritage;
- (b) develop training and research, and share experience and knowledge among the institutions and professional associations concerned;
- (c) encourage universities and other research organizations, both public and private, to ensure preservation of **research data**.





# Technological Change

In the 1990's changes in technology

- Memory becomes cheap
- Proliferation of software and analytical tools

The new century saw the development of:

- IT infrastructure that could pipe huge amounts of information around the globe
- Grid computing, AAA (authentication, authorisation and accounting)
- Growing user expectations
- E-research demands

# E-Research Infrastructure

Change in research methods due to developments in information and communications technology.

- Quick flow of information
- New data visualisation tools
- Focus on connectivity and interoperability between researchers, disciplines and organisations
- Data repositories are now considered part of research infrastructure

# European Strategy Forum on Research Infrastructures



Envisaged in 2000, ESFRI attempts to define and support European Research Infrastructure needs for the next 10-20 years. Projects in the social science covered include

- European Social Survey – funds data collections
- CESSDA – Council of European Social Science Data Archives, integrated catalogue based on NESSTAR
- EROHS – European Research Observatory for the Humanities and Social Sciences – data integration (see <http://www.erohs.org/>)

Recognition that the survey is to a social scientist as a telescope is to an astronomer – vital infrastructure

# CESSDA



## Council of European Social Science Data Archives

- European data sharing
- Supports the development of NESSTAR
- Have an integrated NESSTAR data catalogue
- Australia is considered a CESSDA partner

<http://extweb3.nsd.uib.no/cessda/home.html>

# National Collaborative Research Infrastructure Strategy



- The Australian version of ESFRI
- Only funds hard sciences – except for “Platforms for Collaboration”
- Will develop a National Data Management Infrastructure
- Focus is on funding access and interoperability rather than preservation which will be left to the data producers eg institutional repositories

<http://www.ncris.dest.gov.au/>



# Who Pays for Data Preservation in the UK?

ESRC (Economic and Social Research Council) funds the UK Data Archive

- The UKDA checks data producing grant applications to see if the research is new
- Researchers don't get the final grant payment until their data is deposited
- The UKDA applies for other grants to establish new preservation efforts outside its general funding

# Who Pays for Preservation in the USA?



The ICPSR (Inter-University Consortium for Political and Social Research) has a diverse range of funding, but the main ones

- Membership fees to access data holdings
- Attendance fees for statistical training courses
- Grants for specific preservation projects

# Who Pays for Preservation in the Australia?



- Traditionally the bulk of the funding for ASSDA's activities has come from the Australian National University
- The Australian Research Council can pay for archival research projects
- ACSPRI helps support a Requests Officer to help ACSPRI members locate data
- APAC (Australian Partnership for Advanced Computing) provides IT infrastructure and storage





# Sustainability

- All data archives complain about sustainability – having to compete for funds each year to survive puts pressure on doing cool stuff rather than solid stuff
- In Australia, ARC says data must be archived, but cannot directly financially support preservation work
- NCRIS is designed to support access rather than preservation
- ASSDA can get research and access funded externally – but preservation is only funded by host institutions
- So to increase sustainability, we increased institutional involvement

# ASSDA Advisory Panel

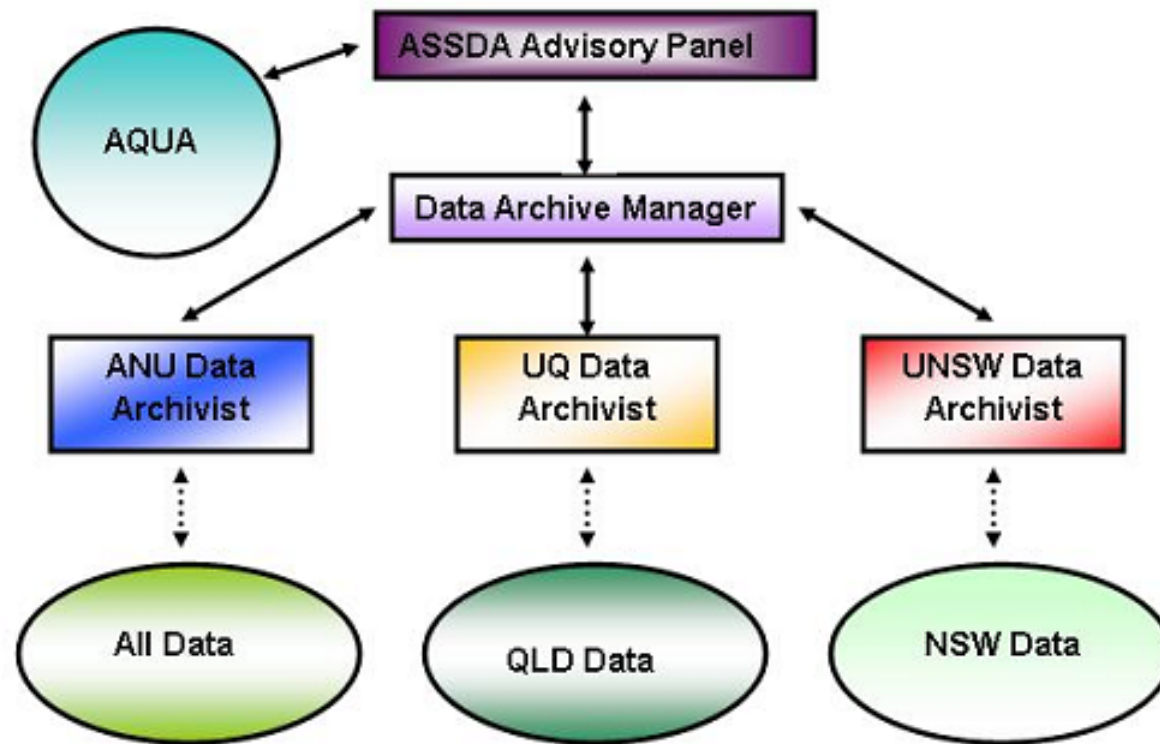


All decisions about the directions of ASSDA developments are made by the ASSDA Advisory Panel



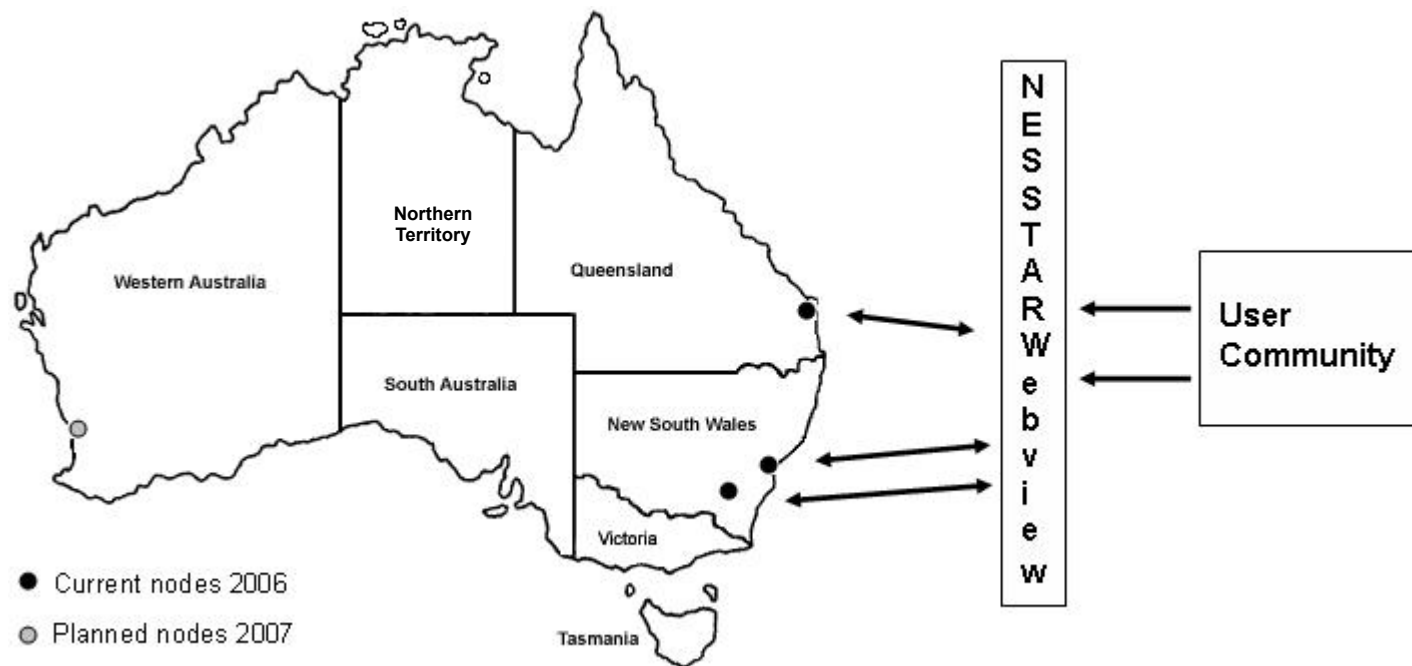
# Management Structure

All decisions about the directions of ASSDA developments are made by the ASSDA Advisory Panel



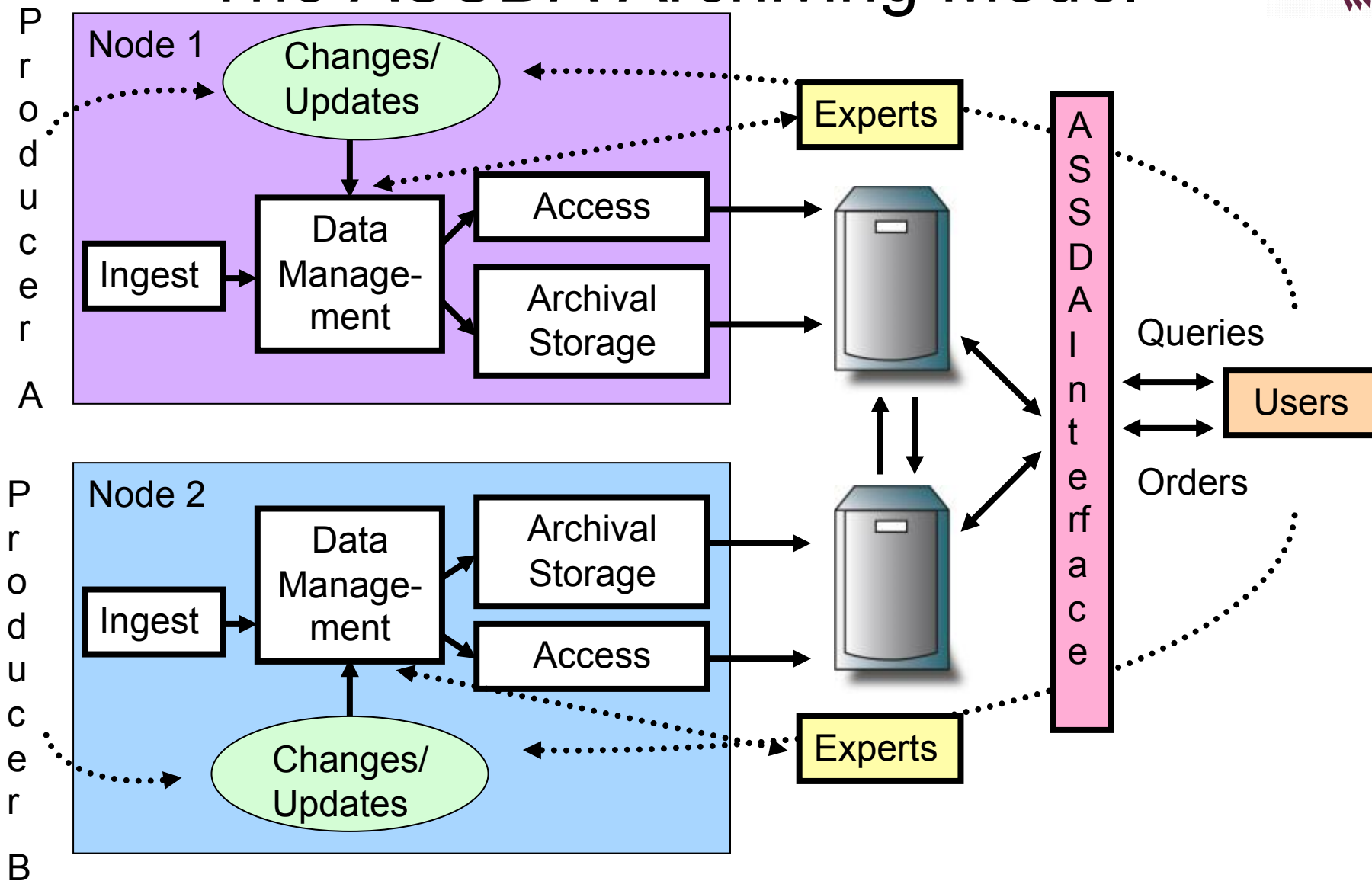
# ASSDA's Current Structure

- In 2007-ANU UQ UWA UNSW
- In the future – UMelb Utas



<http://assda-nesstar.anu.edu.au/webview/index.jsp>

# The ASSDA Archiving Model





# Role of Nodes

- To decentralise data processing, storage and funding
- To give local ownership of their data – rather than sending it off to privilege
- To have active collection efforts at a local level
- To collect a critical mass of expertise to further preservation efforts of difficult data types

# ASSDA-UQ



- General preservation – UQ only data
- AQuA – Australian Qualitative Data Archive
  - Leximancer being developed for privacy
  - Developing a different metadata standard
  - Working closely with Qualidata from the UKDA
  - Contact Andrew Smith on [andrew@humanfactors.uq.edu.au](mailto:andrew@humanfactors.uq.edu.au)

# ASSDA-UNSW



- General preservation from SPRC
- Was meant to specialise in Administrative – no luck
- Now in discussions with HIV Research



# ASSDA-UWA



- General preservation of Western Australian Data
- Specialising in Election data
- Will be preparing data cubes using the Campbell Sharman election data
- Contact A/Prof David Denemark on [denemark@cyllene.uwa.edu.au](mailto:denemark@cyllene.uwa.edu.au)

# ASSDA-ANU



- Server management
- Setting standards
- Managing requests/access
- Projects experimenting with different data types including:
  - Historical and Colonial Census Data Archive
  - The Indigenous Data Archive



# HCCDA

- Historical Census and Colonial Data Archive
- ASSDA is working with Dr Len Smith - Demography, Prof Tim Rowse - History, Stuart Hungerford – ANU SuperComputer and Robyn Cammack - ABS
- ARC Funded

The screenshot shows a web browser window displaying the HCCDA website. The page title is "HCCDA" and the URL is "http://150.203.5.215/assda/cwith/year/1961/volume/8/part/5R/chapter/XX/page/329". The page content includes a search bar, navigation links, and a table titled "9. Density of Population and Occupied Dwellings".

**9. Density of Population and Occupied Dwellings**

The next table shows, for the Metropolitan Urban, Other Urban, and Rural Divisions of each State and Territory, the density of population and of occupied dwellings.

**AREA, POPULATION AND DWELLINGS IN URBAN AND RURAL DIVISIONS, 30th JUNE, 1961**

Particulars	New South Wales	Victoria	Queensland	South Australia	Western Australia	Tasmania	Northern Territory	Australian Capital Territory	Australia
<b>METROPOLITAN URBAN</b>									
Area in Square Miles	671.38	812.17	474.25	161.42	191.88	104.97		43.38	2459.45
Population	2183388	1911895	621550	587957	420133	115932		56448	5897304
Occupied Dwellings	609749	524608	168639	163900	114848	30580		13372	1625696
Population per Square Mile	3252.09	2354.06	1310.60	3642.40	2189.56	1104.43		1301.27	2397.81
Occupied Dwellings per Square Mile	908.20	645.93	355.59	1015.36	598.54	291.32		308.25	661.00
<b>OTHER URBAN</b>									
Area in Square Mile	2128.61	569.82	835.14	192.21	185.11	149.43	59.00		4119.32

The right side of the screenshot shows a preview of the original document page, which includes the title "9. Density of Population and Occupied Dwellings" and the same table as shown in the main content area.



# IDA

- The Indigenous Data Archive
- Pilot in 2007
- Provide a culturally appropriate workspace to allow correct preservation and access to indigenous data
- Capacity building of Indigenous students
- Specialist Advisory panel of Indigenous researchers



# Social Science Infrastructure

- Surveys are social science infrastructure
- Data archives provide this infrastructure in a manner required by their user community
- NESSTAR is a tool which will allow countries like Australia and New Zealand to plug into the international community through CESSDA
- New tools need to be developed which take advantage of grid computing and make data more accessible to a wider research community



# Conclusion

- ASSDA responded to changes in culture and technology and remains a vital resource to the social science community in Australia
- ASSDA hopes to have a long working friendship with SSDASH and NZSSN to allow us both to participate in international initiatives
- ASSDA also hopes that we can learn from and share in the developments led by SSDASH and NZSSN here in New Zealand and apply them to the Australian context