ASSDA: A Trusted Digital Repository or a trusted digital repository?

Sophie Holloway
The Australian Social Science Data Archive
Taking the Shock Out of the Future
31 August - 1 September
Overview

• Introduction to ASSDA
• ASSDA in OAIS terms
• ASSDA’s position of trust in its community
• ASSDA against the TDR Checklist
• Issues of compliance to the TDR
• Some positives and negatives of the TDR system
Introduction to ASSDA

ASSDA was set up in 1981, housed in the RSSS, ANU

It was established by a group of Australian researchers worried about the lack of preservation of Australian Social Science data

ASSDA was modeled on the ICPSR data archive at the University of Michigan and the UK Data Archive at the University of Essex in both structure and standards

In 2006 opened new nodes at UQ and UNSW
Preservation: What we collect

We collect any machine-readable data from the quantitative social sciences from the Australian academic, government and private sectors.

Our most important academic collections include
• The Australian Election Study
• The Australian Study of Social Attitudes
• Opinion Polls from 1943 to present

Our most important government collections include
• The Historical Census Archive for the ABS
• The National Drug Strategy Survey for the Dept of Health

We act as custodians of data from other nations in the Asia Pacific upon request.
Access: Who uses ASSDA

Since the release of ASSDA’s NESSTAR Data Discovery and Distribution System in September 2005, we are averaging around 200 new users each month. Those using the online analysis features are largely students – simple analyses - from smaller universities and overseas. Those using the data download function are mostly postgraduate students or researchers - more complex analyses – mostly from wealthier universities. Some universities run undergraduate research methods classes using ASSDA – up to 300 students.
What is ASSDA’s Information Package

An information package would include:

• a dataset (eg answers to a survey)

• Metadata

• Related materials – any documents used to create the metadata, other useful documents such as questionnaires, codebooks, User’s Guides, Technical Reports

• Administrative Data – communications with copyright holders, records of changes made to the package
SIP – Submission Information Package

Preservation Implications

• Files deposited in proprietary formats
• Software and hardware dependent
• Metadata not created to a standard

Access Implications

• Format deposited in may be unsuitable for use in wider community.
• Difficult to compare data sources when not standardised
AIP – Archival Information Package

Archival version of data converted to text

• The internal structure of the file is standardised to help future migration

• Standardised metadata created to international standard

• All communications with package owner and decisions made stored electronically with package

Problem:

• Archival format not user friendly
DIP – Dissemination Information Package

The package is converted to be NESSTAR compatible. NESSTAR:
• stores one copy of the package in NSDStat format
• converts data to one of nine popular formats at the time of download
• allows all relevant documents to be attached to package, and relevant weblinks in the metadata

The NESSTAR online catalog also supports browsing, searching, viewing data, simple analyses
ASSDA’s DIP
ASSDA’s DIP

Dataset: The Australian Survey of Social Attitudes, 2003

Variable v2 : [v2] Power in Australia: Big business

Pre-Question
Thinking about the amount of power organisations have in Australia today, please say whether you think each of the following should have more power, less power or the same amount.

Literal Question
a. Big business

Values Categories N
1 A lot more power 61 1.5%
2 More power 146 3.6%
3 The same amount of power 1324 33.2%
4 Less power 1001 45.3%
5 A lot less power 636 14.3%
6 Can’t choose 132 3.2%
-2 Skip 0 0.0%
-1 Missing 0 0.0%

Summary statistics
Valid cases 4110
Missing cases 163
This variable is numeric

Copyright © 2004 NESSSTAR - All Rights Reserved - neestar-support@nessstar.com
## Dataset: The Australian Survey of Social Attitudes, 2003

<table>
<thead>
<tr>
<th>Variable Description</th>
<th>Gender</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power in Australia: Big business</td>
<td>Female</td>
<td>Column percentage</td>
</tr>
<tr>
<td>Power in Australia: The Federal government</td>
<td>Male</td>
<td></td>
</tr>
<tr>
<td>Power in Australia: The mass media</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>Power in Australia: Unions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Which social class would you say you belong to?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tension between: Management and workers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tension between: People born in Australia and migrants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tension between: Consumers and corporations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Put self on nobles: Top bottom of society</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Federal government run for big interests or people</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often use for news: Commercial television</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often use for news: ABC and/or SBS television</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often use for news: Radio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often use for news: Newspapers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often use for news: Internet sites</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often use for news: Talkback radio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often use for news: News magazines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relied on MOST for your news and information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreed: Media in Australia effective in keeping govt_info honest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreed: Media in Aust. effective keeping govt_info on their toes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreed: Violence on television encourages violence in society</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreed: Media ownership concentrated among few rich families</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreed: Television main source of entertainment in our house</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use the Internet at Home</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use the Internet at Work</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use the Internet at Other (e.g. University, internet cafe)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c) 2004 The Australian National University. All rights reserved.
**ASSDA’s DIP**

---

**Dataset: The Australian Survey of Social Attitudes, 2003**

**Download**

Please select a data format from the drop-down box. If you wish to download only a subset of the data, click on the "Subset" button. Click on "Download" to start downloading. Please note that you may be asked for a password.

<table>
<thead>
<tr>
<th>Subset Data Format</th>
<th>Download</th>
<th>Subset</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPSS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPSS Portable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCDOSS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statsoft</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stata v8</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>Stata v7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[Variable Description]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[v1] Identification number</td>
</tr>
<tr>
<td>[v2] Power in Australia: Big business</td>
</tr>
<tr>
<td>[v4] Power in Australia: The mass media</td>
</tr>
<tr>
<td>[v5] Power in Australia: Unions</td>
</tr>
<tr>
<td>[v6] Which social class would you say you belong to?</td>
</tr>
<tr>
<td>[v7] Tension between: Management and workers</td>
</tr>
<tr>
<td>[v8] Tension between: People born in Australia and immigrants</td>
</tr>
<tr>
<td>[v9] Tension between: Consumers and corporations</td>
</tr>
<tr>
<td>[v10] Put self on scale: Top-bottom of society</td>
</tr>
<tr>
<td>[v11] Federal government run for big interests or people</td>
</tr>
<tr>
<td>[v12] How often use for news: Commercial television</td>
</tr>
<tr>
<td>[v13] How often use for news: ABC and/or SBS television</td>
</tr>
<tr>
<td>[v14] How often use for news: Radio</td>
</tr>
<tr>
<td>[v15] How often use for news: Newspapers</td>
</tr>
<tr>
<td>[v16] How often use for news: Internet sites</td>
</tr>
<tr>
<td>[v17] How often use for news: Talk-back radio</td>
</tr>
<tr>
<td>[v18] How often use for news: News magazines</td>
</tr>
<tr>
<td>[v19] Relly on MOST for your news and information</td>
</tr>
<tr>
<td>[v20] Agree: Media in Australia effective in keeping govt’s honest</td>
</tr>
<tr>
<td>[v21] Agree: Media in Aust. effective in exposing govt’s on their bias</td>
</tr>
<tr>
<td>[v22] Agree: Violence on televiewing encourages violence in society</td>
</tr>
<tr>
<td>[v23] Agree: Media ownership concentrated among few rich families</td>
</tr>
<tr>
<td>[v24] Agree: Television main source of entertainment in our house</td>
</tr>
<tr>
<td>[v25] Use the Internet at: Home</td>
</tr>
<tr>
<td>[v26] Use the Internet at: Work</td>
</tr>
<tr>
<td>[v27] Use the Internet at: Other (i.e., university, internet cafe)</td>
</tr>
</tbody>
</table>
The ASSDA Model

- Changes/Updates
- AIP
- Archival Storage
- Access
- Queries
- Orders

Nodes:
- Producers
- Consumers
- SIP
- DIP

Arrows:
- Ingest
- Data Management

Graph:
- PRODUCER to INGEST
- INGEST to DATA MANAGEMENT
- DATA MANAGEMENT to ARCHIVAL STORAGE
- DATA MANAGEMENT to ACCESS
- ARCHIVAL STORAGE to AIP
- AIP to CHANGES/UPDATES
- CHANGES/UPDATES to DATA MANAGEMENT
- DATA MANAGEMENT to INGEST
- INGEST to PRODUCER
- CHANGES/UPDATES to AIP
- AIP to queries
- Queries to orders
- Orders to DIP
- DIP to CONSUMER
- CONSUMER to INGEST
- INGEST to PRODUCER
- CONSUMER to AIP
- AIP to CHANGES/UPDATES
Designated Community

Members of ACSPRI (Australian Consortium for Social and Political Research Incorporated)

- Australian Tertiary Institutions
- Australian government departments with a strong focus on social research

Other groups with an interest include

- Other social researchers
- Members of IFDO (International Federation of Data Organisations)
Trust in ASSDA

ASSDA is highly regarded and trusted by its designated community

• ARC Grant contracts stipulate that any recipients of ARC funds which produce social science data must deposit with ASSDA

• Many groups deposit as a standard and final step of any appropriate data gathering project

• We don’t pay for any data

• We are invited to sit on government projects

• We work with other countries to establish archives

But do we measure up?
ASSDA as a TDR

A. The Organisation ★★★
B. Repository Functions, Processes & Procedures ★★★★★★
C. Designated Community and Usability of Information ★★★★★★
D. Technologies and Technical Infrastructure ★★★★★★

Let’s look at ASSDA’s strengths first…
B. Repository Functions, Processes & Procedures

- ASSDA clearly defines its SIP to its designated community
- ASSDA defines its AIP, and records all processes occurring between the SIP and the AIP
- ASSDA creates metadata, stores it with the AIP, and manages all access in accordance with deposit agreements
C. Designated Community and Usability of Information

- ASSDA exists within its designated community
- Staff are trained within the social sciences, and continue to be involved in research projects
- Uses the metadata standard DDI – developed specifically for social scientists
- Developments, directions and standards are set by community leaders and tested within the community
D. Technologies and Technical Infrastructure

- ASSDA uses well supported appropriate software and hardware, with daily back-ups and synchronised storage at different locations.
- Tests changes, has a disaster recovery plan
A. The Organisation

This is the part of the audit checklist we would most struggle with and it all comes down to financial sustainability

Our Strengths:

• A clear mission statement
• Highly skilled staff
• Strong rights managements policy and procedures
A. The Organisation

Our weaknesses:

- Not enough staff
- No time to be able to regularly review and update policies and procedures – wait and watch while others adopt new methods, and learn from their mistakes
- Financial sustainability not great – can’t charge users enough to cover costs – social sciences excluded from many funding opportunities in favour of pure sciences and ICT – social science money tends to go on original collections rather than preservation.
# ASSDA’s Funding Sources

<table>
<thead>
<tr>
<th></th>
<th>Preservation</th>
<th>Access</th>
<th>R&amp;D</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="#">ANU</a></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><a href="#">UNSW</a></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td><a href="#">University of Queensland</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="#">Australian Government</a></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td><a href="#">Australian Research Council</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="#">ACSPRI</a></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
ASSDA’s Funding

• ARC only funds Research and Development, they don’t provide ongoing support. A 2 year cycle.
• The rest of the designated community is only willing to contribute to Access – easy to sell
• But who pays for preservation?!?
• Existing within an academic department means that any financial changes within the University directly effects the archive (academic vs general staff ratios)
• Until there is ongoing support for centralised recurrent funding for the ongoing preservation ASSDA will not be able to meet Section A.
Implications of the TDR – The Negatives

• We may not go well on the checklist because of restrictions in the Australian social science community.
• If we don’t go well, will people still trust us?
• How will that effect our international standing?
Implications of the TDR – The Positives

• The TDR Checklist gives repositories a clear idea of what they should aspire to
• Attaining these goals could be used as the basis for funding applications
Conclusion

• ASSDA has been in business for over 25 years
• We occupy a position of trust within our designated community
• We are trusted as a permanent fixture, with strong standards and policies, and we work hard to provide valuable infrastructure to its designated community
• While our designated community remains the way it is, we may never get perfect marks as a Trusted Digital Repository, I think we will always work hard to remain a trusted digital repository