



The Role of Government Data Collectors: Opportunities & Responsibilities for Research

*Long-lived Collections:
the future of Australia's research data*

11 September 2007

National Library, Canberra

statistics for informed

decision making

steve.matheson@abs.gov.au

National Statistical Service Leadership Branch

Topics

- Long-term data storage and access in the government sector
- Role of government agencies
- Special responsibilities?
- Opportunities to aggregate or share data?
- Getting the best return on investment on data/information assets

ABS and the National Statistical Service (NSS)

- ABS – just part of picture
- Increasing importance of other sources of data/information
 - administrative data
 - all records potentially available
 - “higher resolution”
 - social research data
 - scientific data
- ABS – nearly all data products free via www.abs.gov.au

ABS and the National Statistical Service (NSS)

- ABS coordination role
 - data/metadata standards
 - data collection methods
 - data quality
 - facilitate the effective use of government data/ information
- ABS – Data for Science Working Group
 - arose from active role in visibility, accessibility, interoperability space

Fundamental Questions?

- What's out there? [Visibility]
- How can I find it? [Search]
- Is the information resource in reasonable shape?
[Information Asset Management]
- Can I access it? [Access Conditions / Controls]

Data Domains?

- Survey/census data
- Administrative data
 - births, deaths
 - Medicare, pharmaceutical benefits
 - other individual/family benefits
- Scientific data
 - Bureau of Meteorology
 - Geoscience Australia
 - CSIRO
 - other

Contrast: Australia/USA

- Australia
 - data/information produced by government agencies belongs to the crown
 - visibility/accessibility at discretion of crown
- USA
 - data/information produced by government agencies belongs to the people
 - agency obligations to people

Information Policy for Australia?

- Existing elements– legislation, other
- Missing elements:
 - overarching principles
 - mechanism to address market failure for significant data/information holdings
 - interoperability between domains
 - data management, accessibility obligations of CEOs
 - enablers: IT, legal, cultural
 - national data management agency?

Fundamental Principles

- Data/information
- *managed as an asset*
- *exposed as broadly as practicable*
- *used responsibly*
 - *conditions of use*
 - *privacy principles, applicable laws*
- *quality of data/information clearly described*
- Work collaboratively, avoid duplication of effort

Getting the Ball Rolling?

- Most agencies are interested in everyone else getting their data organised, visible, accessible first!
- Ways of getting agencies to contribute to the establishment of a controlled visibility/ access, distributed network of government data/ information resources?

Getting the Ball Rolling?

- ABS & National Data Network
 - distributed network infrastructure
 - controlled visibility and access of information resources
 - machine-machine not just person-machine interactions
 - open source, free-of-charge
 - active in all areas impacting on visibility/access

ABS & NDN: Progress-to-date

- Distributed network software infrastructure (open source) has been developed – over 25 staff years development to date
- NDN Demonstration Phase August 2006 to June 2007 complete
- NSN Pilot Phase July 2007 to October 2007 underway
- Children and Youth Statistical Portal pilot project
- Collaboration with State governments, in particular QLD and VIC

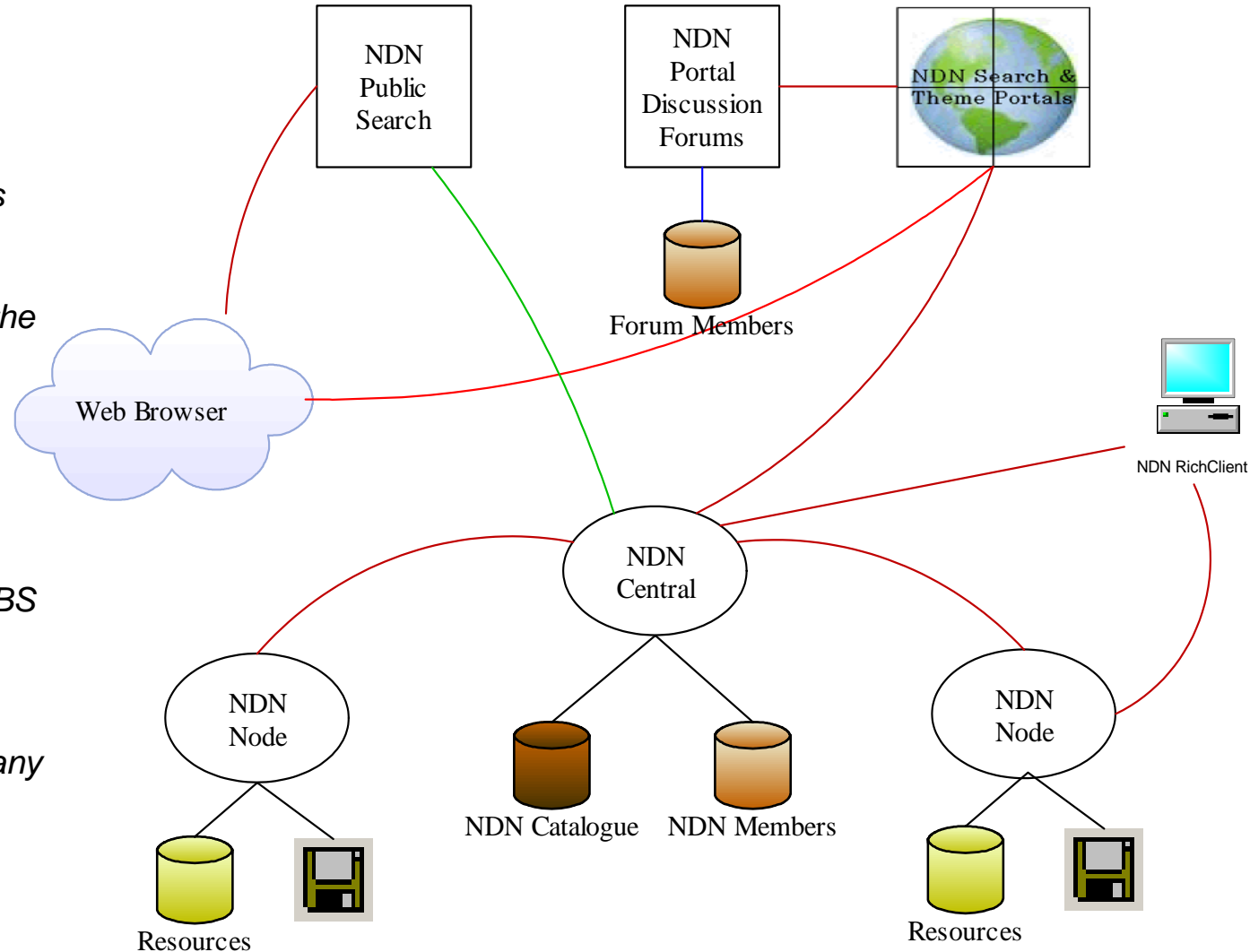
NDN Architecture – using open source technology and standards

NDN content




(information resources*):

- documents, spreadsheets
- database files, presentations
- PDFs, MP3s
- links to resources found on the Web
- pod casts
- geospatial services
- statistical services
- SOAP Web services (e.g. ABS address coder)
- restful Web services

* any type of file and almost any type of 'service'



NDN Current Collaborations

- Collaborative development agreement is in place with the US Census Bureau – DataFerrett 
- Collaboration on Shibboleth Federated Identity Management – MAMS project (Macquarie Uni. MELCOE)  Shibboleth.
- Collaboration on Creation Commons (digital licence management) 

NSO Collaboration Opportunities - What is next?

Your collaboration? Some possibilities:

- *Shibboleth – Identity Management*
- *Geospatial Services*
- *Federated Searching*
- *Statistical Web Services*
- *RSS Feeds*
- *Tagging of Information Resources*
- *DataFerrett integration*
- *Privacy Preserving Analytics*
- *Creative Commons*
- *Automated workflows*
- *Warehousing of Information Resources*
- *Connections to other network*

Owner/Custodian Services

Researcher/Analyst Services



Questions?