Australian Government Gazettes

Constructing a Textual Reference Dataset for Research Reuse
tinker.edu.au/data/available-datasets/
github.com/stevecassidy/hassdevl-samples/
www.aarnet.edu.au/network-and-services/cloud-
services-applications/cloudstor
Curatorial and technical analysis

**Custodial**
- Custodial arrangements
- Collection access

**Research**
- Research methodologies
- “Machine ready” data

---

**Australian Cultural Data Commons**
- Data infrastructure
- Data transfer methods
- Data processing and enhancement
Three people (and platforms)

- Data custodian – Julia Hickie via Trove (National Library of Australia)
- Researcher – A/Prof Steve Cassidy via Alveo VL (Macquarie University)
- Analyst – Ingrid Mason via CloudStor (AARNet)
Julia Hickie
@TroveAustralia
A/Dir Trove Data, Discovery & Delivery
National Library of Australia

The Australian Government Gazettes
1832-1968
A/Prof Steve Cassidy
@SteveCassidyMQ
Academic in computational linguistics
Alveo virtual lab @ Macquarie University
Ingrid Mason
@1n9r1d
Deployment Strategist, eResearch
AARNet

Data Commons Considerations

• Corpus arrangement:
  • Consistency (readability)
  • Flexible access (systematic)
  • Context (sense-making)

• Data transfer techniques:
  • Query
  • Packaged
  • M2M

• Referenceable data:
  • Ongoing (updated)
  • Snapshot (static)
  • Subset (derived)
  • Enhanced (derived)