



COCOSDA Report: Oceania

Steve Cassidy

Summary of Activity

- HCSNet: ARC Research Network in Human Communications Sciences
- eResearch Infrastructure
- ANDOSL: Attempts to release a 16 yr old collection
- Proposal: An Audio-Video Speech Corpus of Australian English

HCSNet

- Australian Research Council funded Network
- Supports collaboration and activity across Human Communication Sciences
 - NLP, Phonetics, Acoustics, Music, Sociolinguistics...
- Priority Areas:
 - Speech
 - Human Effective Interfaces
 - Next Generation Search Technology
 - Human Communication Disorders
 - Perception and Action

eResearch Infrastructure

- 2005 ARC eResearch funding for a number of LR projects
- Eg. DADA-HCS provides infrastructure for sharing LRs and collaborative annotation
- A number of National projects building data archives for science and humanities
- Australian Access Federation: authentication and authorisation services
- 2008 NSW Intersect eResearch centre may support further development of DADA-HCS

ANDOSL

- Significant collection of Australian English circa 1992
- Many speakers, studio quality read speech, passages, map task
- Currently held by the original partner universities
- Australasian Speech Science and Technology Association
 - Currently negotiating release of data for redistribution
 - Will host the collection and will accept new annotations on the data
 - May be for ‘research purposes only’

Proposed ANDOSL2

- 27 collaborators over 13 Institutions
- \$1.3M for 1 year (restriction of the funding scheme)
- Establish **Standard Speech Science Infrastructure Black Box**
 - Audio, video recording hardware platform
 - Suitable for use in remote locations
- Team of Research Assistants perform recording and base annotation
- Central technical team provides support for annotation
- Central server to publish data as it is completed
- Coverage of geographical variation, migrant accents, disordered speech