

# Data and the Humanities

Professor Graeme Turner,  
President, Australian Academy of  
the Humanities

# Data and the humanities

- What place does data play in the humanities?
- What forms does it take?
- What is the current state of the relevant research infrastructure?
- What needs to be done?

# Common assumptions about humanities research

## Assumption 1

- That the humanities do not use data; they just use libraries. (And that they are technophobic!)

## Assumption 2

- That humanities research is solo, non-collaborative and therefore does not require networking

## Assumption 3

- That electronic research infrastructure does not need to take account of the humanities' specific requirements.

# Two responses to these assumptions

## ***Humanities Technologies: research methods and ICT use by humanities researchers***

LASP project, AHA, led by John Byron and Sarah Howard

Survey into the current practices and future requirements of humanities researchers in Australia. Aimed at informing the planning and provision of critical technological infrastructure and training for humanities researchers.

# Some interesting findings so far

- Only 53% of researchers were working as sole researchers on their first project; 42% on their second project, and 35% on their third project.
- Interdisciplinarity was common with 12% of first projects, 15% of second, and 17% of third projects falling into this category.
- They made significant use of e-resources and particularly of those primary texts that were available electronically (Primary texts: books, journals, corpora, manuscripts; Secondary texts: critiques, criticism, reviews, editions).
- While non-electronic primary resources remained important, for researchers with a profile of 6-10 years and 11-20 years experience, electronic primary texts were more important than all other types of e-resources, with secondary texts and databases, indexes and catalogues and archives next in importance.

# Humanities researchers and the creation of electronic resources

Uniform results here across disciplines:

- Texts (articles, chapters, books) by far the most common form of e-resource being created by the survey respondents, following by databases and datasets, and digitising their own resources for own use in research.

Difficulties cited included

- Lack of expertise
- Licensing and copyright restrictions
- Quality and reliability of resources (on the Web and elsewhere); and
- Shortage of discipline specific software

# Importantly.....

- Large majority of respondents agreed that they would change their research practices to include new research technologies if more arts and humanities ICT infrastructure was available.
- When asked to choose between options, this is how they panned out:
  - More resources created, ie digitisation of primary resources used in research: 59%
  - Better tools to help with search, retrieval and processing: 41%

# What needs to be done

At present very few of our basic research resources are available electronically – very few literary texts, very few historical documents, very few newspapers, and so on.

The decision to regard textual material such as these as content rather than as infrastructure has been deleterious and it is going to take a major program of digitisation to correct it.

The AHA has presented the case to the Minister for such a program and as a result DEST is funding a small scoping study of what that would involve.

# *Digitisation of national resources for HASS research*

- This project is a scoping study aimed at providing advice to DEST on the likely shape of a digitisation program for research materials in the humanities and social science.
- Located in the AHA with researcher Phoebe Garrett

# International parallels

- UK – AHRC ICT Strategy programme 2005
- US- ASCLS commission on cyberinfrastructure for HASS  
National Endowment for the Humanities  
Division of Preservation and Access fund.
- China –Century Journals Project

# Study is just beginning...

This project will survey Australian academics to discover what their preferred candidates for digitisation would be so that we can present a prioritized list of materials to DEST for consideration as the first tranche of a major digitisation program for national research materials for the humanities and social sciences.