THETA

Create, connect, consume - innovating today for tomorrow.
Australian National Data Service

Our Vision: More valuable data for Australian research

ANDS is enabling the transformation of:

Data that are: \[\rightarrow\] Structured Collections that are:

- Unmanaged \[\rightarrow\] Managed
-Disconnected \[\rightarrow\] Connected
- Invisible \[\rightarrow\] Findable
-Single-use \[\rightarrow\] Reusable
How will (digital) humanities researchers in the future use cultural data?
Connected topics

- Context
- Current
- Future
- Transition
Context

● (Digital) humanities research
● Cultural data
● GLAM data
(Digital) Humanities Research
Cultural Data

humanities, arts, social science, science
Cultural collections on nodes

**Intersect:** A History of Aboriginal Sydney, CSU Regional Archives, Cultural Collections UNEW, FAIMS Repository, Hidden Testimony: musical experience and memory of Jewish Holocaust survivors, HCS VLab Corpora (PARADISEC, AusTalk, AVOZES, Colloquial Jakartan Indonesian Corpus), Multicultural Research Library, Stage on Screen, Sydney Playground Research, Talking Ngan’gi, UNE Archives & Heritage Centre.

**eRSA:** AusStage.

**QCIF:** Historical coastlines (community perspectives) : manuscript and images archive, Anthropology Museum Digital Storage, Solomons Islands History, Marxan Software archive, HCS Vlab Corpora (AusNC).

**VicNode:** AURIN, History of Adoption: Stories of Australian Adoption, Twitter.
RDS: Culture & Communities

- Statistical data, manuscripts, documents, artefacts and audio-visual recordings
- Diverse array of repositories … many of which are unconnected

National Data Collection
RDS: Culture & Communities

• Enable researchers to assemble, combine and analyse data sets at a scale not previously possible, to produce holistic answers to complex questions...

Strategic Roadmap 2011
GLAM Data

galleries, libraries, archives, museums
Cultural collections in Australia

- National GLAM collections
- State and territory GLAM collections
- University GLAM collections
Nexus

- Significant proportion of cultural data is in the public sector in cultural heritage institutions
- Research community resides mostly in the research sector and their scholarly outputs often end up in cultural heritage collections
Current

- (Digital) humanities research
- Use of cultural data
- Cultural data types
State of play

- Analysis methods: qual/quantitative
- Old and new research practices
- New skills are needed for new practices, e.g. using tools or working in teams
- Data and technology intensive
Current (digital) humanities research

- new methods
- new scholarly practices
- new peer communities
- new patterns (data)
- new demand for content = data
New data requirements
Discovery & access tools

Catalogue, finding aid & data service
New access requirements to bulk cultural (GLAM) data
New methods

Computation, high resolution, bulk data
New ways of looking

Close and distant reading...
In common

Kenderdine Manovich Sherratt Whitelaw
GLAM data
researcher relationship with cultural institution
Bulk data
computation + arrangement + distance
Kenderdine
Immersion and embodiment
Kenderdine

Museum Victoria
Europeana
Hong Kong Maritime Museum
Dunhuang Academy
Comparative visualization: Betty Boop cartoon (left) vs. music videos (center and right)
Manovich

Broad view and pattern analysis
Manovich

ArtStor

Whitehouse.gov

Instagram
TroveNewsBot

@TroveNewsBot
Built with the knowledge of 100 million newspaper articles and the awesome power of the Trove API.
github.com/wragge/trovenews...

https://twitter.com/TroveNewsBot
Sherratt

Opening up the archive
Sherratt

National Library of Australia
National Archives of Australia
Flickr
Whitelaw

Generous interfaces
Customs House, Brisbane
stencil, 1965

Mike Kitching, Jollyn's Hungry Horse Art Gallery (1987)
KITCHING, Michael - 1987
Whitelaw

National Archives of Australia
National Gallery of Australia
Manly Library (Sydney, NSW)
DH2015
Current data use in digital humanities
“The Muninn Project is a multi-disciplinary, multi-national, academic research project investigating millions of records pertaining to the First World War in archives around the world.”
“This combined method for cross-document coreference allows for the emergence of narratives that go beyond the boundaries of one interview. Using a test corpus of 511 World Trade Center Task Force Interviews with first responders, this technique reveals the stories of some who did not survive."
Mapping the Dutch Cultural Industry. C van den Heuvel et al.

“The project started with the integration of three complementary, but heterogeneous (meta)datasets.”

Biographical (art history) reference works
Players in the cultural industry
Artists and scholars
“Medieval scripts are a challenge to historical analysis, as for describing and representing the graphical evidence, analyzing and clustering letter forms and their features through Computer Vision and analyzing historical phenomena.”
Genetic Criticism. D Van Hulle et al.

“research focus is the study of modern manuscripts and writing processes, especially comparative genetic criticism, digital scholarly editing and the analysis of manuscripts by authors such as Samuel Beckett, James Joyce, Thomas Mann, Marcel Proust, Charles Darwin, Willem Elsschot.”
“...create high-quality digital reconstructions of the mural paintings and related art and texts associated with the several hundred Buddhist cave shrines in Dunhuang, China, a uniquely important cultural crossroads on the ancient Silk Route in the Gobi Desert.”
What is an element? An element is a substance that cannot be broken down into simpler substances by chemical means. An element is made up of a fixed proportion of one type of atom. What is a compound? A compound is any substance made up of two or more elements that are chemically combined. A compound is made up of a fixed proportion of its elements. How can you tell the difference between an element and a compound? A compound can be separated into its elements through chemical means. Electrolysis is a process used to separate the elements of a compound. How do you separate the elements of a compound? Electrolysis is a process that uses electricity to separate the elements of a compound.
Cultural data types
Data Types

What sort of data supports humanities research?
Transcription system:
Images are delivered to data entry page; they float above text entry boxes; image and text boxes scroll independently

<table>
<thead>
<tr>
<th>register_number</th>
<th>1</th>
<th>yes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>trial_place</td>
<td>2</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>court</td>
<td>3</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Def_surname</td>
<td>4</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Def_firname</td>
<td>5</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Def_othername</td>
<td>6</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Def_sex</td>
<td>7</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>
It rains a lot in Portland.
Future

- (Digital) humanities research
- Use of cultural data
Future humanities research is..

Inextricably tied to the capacity and future of cultural collecting

- bridging and dovetailing between research & collection infrastructures
- research & collecting practices change (data and technology intensive)
Future humanities research

Reliance

- soft and hard infrastructure RDM
- data is digital, available & collected
- data seeking & citing (literacy) skills
- tool know-how & skills
Let’s dream a little..
(Near) future cultural data use
Imagine..

Farah
3 virtual labs
60 hours visualisation
visualisation expert
travel posters

Sydney AUSTRALIA

How did Australia project itself to the world as a travel destination over time?
Imagine..

Carrie
digital repository
digitisation & OCR
ontology expert
fictional works
Imagine...

Tom publishing platform software output project manager multimedia

Cross-Platform
Scholarly publications are changing in their form.
Imagine..

Colin & Mick
text encoding
big data
HPC expert
newspapers

We want to look at the changes to advertising over time in Australian newspapers and align that with changes in law and social convention.
Transition

- Keep up with the changes
- Rethink how research support is provided
We know

- Humanities research is changing and increasingly data and tech intensive.
- What & where cultural data is and how humanities researchers are using it.
We know

- Undigitised data (cultural material) is like “dark matter” (Maltby).
- Format shifting cultural heritage material to digital is a challenge to overcome.
We don’t know

- Where the gaps in data and technology support are.
We can do

• Talk to humanities researchers about what cultural material they would like access to (as data).

• Assist humanities researchers in preparing a case for digitisation, even better help manage and fund that!
Take action to build capacity

- Learn about and instruct on using APIs and mashing or mixing data
- Provide resources, training and support or guidance around, e.g. data analysis, data management, data visualisation.
Boom Tish! Thank You :)

ingrid.mason@intersect.org.au
ingrid.mason@ands.org.au
@1n9r1d