Libraries and the Digital Humanities: Partnership, Collaboration and Shared Agendas

Deb Verhoeven (@bestqualitycrab)
Toby Burrows
ADVANCING DIGITAL HUMANITIES
Research, Methods, Theory

Edited by Paul Longley Arthur and Katherine Bode

DH2015
Global Digital Humanities

Conference Convenor: Professor Paul Arthur
(University of Western Sydney)

PC Chair: Professor Deb Verhoeven
(Deakin University, Australia)

Sydney, Australia

This will be the first time that the annual conference of the Alliance of Digital Humanities Organizations (ADHO) is held outside of Europe and North America in its 26-year history. The theme of Global Digital Humanities acknowledges the field’s expansion worldwide across disciplines, cultures and languages.

29 June – 3 July 2015

DH2015 is hosted by the University of Western Sydney’s Digital Humanities Research Group, a leader in collaborative digital humanities in the Asia-Pacific region. The conference is held in partnership with the State Library of New South Wales and in collaboration with the third International Linked Open Data in Libraries, Archives, and Museums Summit.

www.dh2015.org
Rich Connectivity in the Humanities

- **Data**: Disconnected dots
- **Information**: Some connected dots
- **Knowledge**: More interconnected dots
- **Humanities Knowledge**: Highly interconnected, vibrant network
HuNI is...

- **Big** – \(17,605,374\) + entities and growing
- **Multi-disciplinary** and cross-disciplinary – 31 data sources
- **Changing humanities research:**
  - Working with data on a much larger scale
  - Breaking down disciplinary boundaries around data
  - Promoting data sharing and data re-use
  - Encouraging collaboration to enrich data
The HuNI Partnership

Consortium led by Deakin University

- Cultural data providers (10)
- Humanities software developer (1)
- eResearch organisations (2) – lead development agencies – VeRSI and Intersect

Substantial investment:

- Initial NeCTAR grant of A$1.33M
- Partner contributions of A$480,000
- Partner in-kind contributions amounting to >A$1M
Harvest

Data is harvested from partner sites daily via HTTP either using a HuNi’s custom “simple XML” format or the DAI-PMH protocol. The harvest code is custom python and bash scripts.

Pre-process

Where necessary data is pre-processed to ensure it can be properly transformed.

Transform

Custom python code and XSLT templates transform harvested data into the standard HuNi ontology ready for Indexing by Solr.

Index

Documents are submitted to a Solr instance for indexing. The result is a body of indexed documents made up of the most recently harvested versions. This can be quickly searched via an HTTP interface.
# DATA SOURCES

<table>
<thead>
<tr>
<th>Custodian or Owner</th>
<th>Dataset</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian National University</td>
<td>Australian Dictionary of Biography (ADB)</td>
<td>Biography</td>
</tr>
<tr>
<td></td>
<td>Obituaries Australia</td>
<td>Biography</td>
</tr>
<tr>
<td>AIATSIS</td>
<td>AUSTLANG</td>
<td>Linguistic</td>
</tr>
<tr>
<td></td>
<td>Mura</td>
<td>Language</td>
</tr>
<tr>
<td>Consortium led by Deakin University</td>
<td>CAARP</td>
<td>Cinema</td>
</tr>
<tr>
<td>Consortium led by Dictionary of Sydney</td>
<td>Dictionary of Sydney</td>
<td>History, Geography</td>
</tr>
<tr>
<td>Trust</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consortium led by Flinders University</td>
<td>AusStage</td>
<td>Performance</td>
</tr>
<tr>
<td></td>
<td>Find and Connect Australia (8 datasets)</td>
<td>Child Welfare</td>
</tr>
<tr>
<td></td>
<td>Australian Women’s Register</td>
<td>Biography</td>
</tr>
<tr>
<td>Consortium led by University of</td>
<td>eMelbourne: the Encyclopedia of Melbourne</td>
<td>Melbourne</td>
</tr>
<tr>
<td>Melbourne</td>
<td>eGold: Electronic Encyclopedia of Gold in Australia</td>
<td>Gold Mining</td>
</tr>
<tr>
<td>Consortium led by University of</td>
<td>Design and Art Australia Online (DAAO)</td>
<td>Biography</td>
</tr>
<tr>
<td>New South Wales</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consortium led by University of</td>
<td>AustLit</td>
<td>Literature</td>
</tr>
<tr>
<td>Queensland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consortium led by University of Sydney</td>
<td>PARADISEC</td>
<td>Linguistics</td>
</tr>
<tr>
<td>Deakin University</td>
<td>BONZA</td>
<td>Cinema and TV</td>
</tr>
<tr>
<td>Macquarie University</td>
<td>Media Archives Project</td>
<td>Media Industry</td>
</tr>
<tr>
<td></td>
<td>Australian Media History Database</td>
<td>Media Industry</td>
</tr>
<tr>
<td>RMIT University</td>
<td>Circus Oz Living Archive Video Collection</td>
<td>Circus</td>
</tr>
<tr>
<td></td>
<td>Australian Film Institute Research Collection</td>
<td>Cinema and TV</td>
</tr>
<tr>
<td></td>
<td>Australian Film Institute Research Documentation</td>
<td>Documentation</td>
</tr>
<tr>
<td>University of Melbourne</td>
<td>Encyclopedia of Australian Science</td>
<td>Biography</td>
</tr>
<tr>
<td></td>
<td>Saulwick Polls</td>
<td>Social Science,</td>
</tr>
<tr>
<td></td>
<td>Wallaby Club</td>
<td>Politics</td>
</tr>
<tr>
<td>National Library of Australia</td>
<td>Trove Digitised Newspapers</td>
<td>Newspaper</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Documentation</td>
</tr>
</tbody>
</table>
Users of the HuNI virtual laboratory can...

- **Discover and explore** across the aggregated data
- **Make connections** and create “socially-linked” data
- **Save and share** their data and their findings
- **Curate and import** their own data
Socially Linked Data

COLLECTION: Australian Fashion 1850-1950

Details

Created by: Michelle Smith

Links

Australian Fashion 1850-1950 is linked to the following:

<table>
<thead>
<tr>
<th>Record name</th>
<th>Relationship to selected</th>
<th>Relationship from selected</th>
<th>Who made this link?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nancy Greene</td>
<td>contains</td>
<td>contained in</td>
<td>Michelle Smith</td>
</tr>
</tbody>
</table>
SERENDIPITY: The X-factor of humanities research
NEXT STEPS

• 1. Improving the internal connectedness of HuNI using Neo4J and ingesting new internal cross links from existing data sources

• 2. Improving the external connectedness of HuNI by linking HuNI to the Linked Data Cloud and exposing HuNI entities as Linked Open Data with persistent identifiers

• 3. Embedding HuNI in the communities and culture node of RDSI
The database stores user profile information, links between documents, collection lists, and associated metadata.

The public Solr document index allows searching for and retrieving indexed documents via an HTTP API initiated by the client. Documents are sent to the client as JSON objects.

Graph Database

Links between documents and their metadata stored in a graph database to enable fast and expressive searching on records and their link properties.

HTTP

Proxy Server (Nginx)

The proxy server accepts all internet facing requests and delegates them to the appropriate service. HTTP requests are redirected to HTTPS.

HTTP

TCP/IP DB API

Static HTML (Nginx)

HuNI VL Restful API (Perl & Dancer)

User profile information, links, and collections are managed via an API implemented in this component. Data is returned as JSON objects. This layer also enforces access restrictions.

The HuNI project is funded by Nectar, led by Deakin University in Melbourne and is a partnership between VeRSI, Intersect, the University of Western Australia, the Australian National University, RMIT University, the University of Queensland, the University of Melbourne, the University of Sydney, the University of New South Wales, Flinders University of South Australia, AIATSIS.
You say it very well, better than I will say it: we are condemned to live together.
Albert Camus, *Algerian Chronicles*

Our understanding (of the meaning of Being) is an understanding that we share understanding between us and, at the same time, because we share understanding between us: between us all, simultaneously—all the dead and the living, and all beings.
Jean-Luc Nancy, *Being Singular Plural*

We didn’t build our bridges simply to avoid walking on water. Nothing so obvious.
Jeanette Winterson, *The Passion*


THE DIGITAL HUMANITIES

THE FUTURE IS MEOW

memegenerator.net
Harvest
Data is harvested from partner sites daily via HTTP either using a HuNI's custom "simple XML" format or the OAI-PMH protocol. The harvest code is custom python and bash scripts.

Pre-process
Where necessary data is pre-processed to ensure it can be properly transformed.

Transform
Custom python code and XSLT templates transform harvested data into the standard HuNI ontology ready for indexing by Solr.

Harvest
Documents are submitted to a Solr instance for indexing. The result is a body of indexed documents made up of the most recently harvested versions. This can be quickly searched via an HTTP interface.

Linking API
Links between documents are created programatically via the linking API.