Integrating Research in the Humanities: The HuNI Virtual Laboratory

Professor Deb Verhoeven (Deakin University)
Dr Toby Burrows (University of Western Australia)

http://huni.net.au/
The humanities and creative arts cover a wide variety of disciplines, from archaeology and linguistics to film and media.

Existing discipline-based structures and organizations encourage specialization, and discourage collaboration between humanities disciplines.

The same approach has also resulted in a proliferation of discipline-specific digital resources and datasets.

This discipline-centred specialization tends to exclude the wider public and the staff of cultural institutions.

And yet: the “big questions” in the humanities require an interdisciplinary approach and an integrated pool of data.

Working with disciplines outside the humanities also requires an interdisciplinary approach and an integrated pool of humanities data to draw on.
HuNI: Filling the gaps

• Humanities researchers need to be able to work with cultural datasets more efficiently and effectively, and on a larger scale: *HuNI is aggregating datasets from disciplines across the humanities and creative arts*

• Research data in the humanities need to be made available for many types of users – researchers, the curators of cultural datasets, cultural institutions, and the community: *HuNI is providing an environment where many different types of users can work together*

• Researchers need to be able to re-use data and share the results of their work: *HuNI is building a Virtual Laboratory with a suite of tools for working with and sharing data*

• Humanities data need to be made available in formats which can support innovative methodologies and techniques like network analysis and visualization: *HuNI is building a Linked Data service which provides data in a variety of reusable formats*
HuNI: Implementation

• HuNI (Humanities Networked Infrastructure) is one of the Virtual Laboratories funded by the National eResearch Collaboration Tools and Resources (NeCTAR) project

• HuNI is being built by a national consortium led by Deakin University:
  ❖ Cultural data providers (10) – project co-operators
  ❖ Humanities software developer (1) – project co-developers
  ❖ eResearch organisations (2) – lead development agencies – VeRSI and Intersect
HuNI: Key Elements

1. **Integrating data**: bringing together 28 individual datasets across the humanities and creative arts

2. **Integrating meaning**: aligning the meaning and the semantic content of the discipline-oriented digital resources

3. **Integrating the research environment**: building a generic set of tools for users from a range of disciplines

4. **Implementation**: bringing together dataset custodians, researchers, and technical experts; developing a consortial approach; arriving at a common definition of project methods
Integrating Data

- 28 Australian datasets are being harvested for integration into HuNI
- The partner sites expose their data as XML feeds
- HuNI gateway components are deployed on the NeCTAR Research Cloud.
- HuNI harvests the XML feeds and transforms them for ingestion into two data aggregates: a Solr search server and a Jena RDF Triple Store.

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
HuNI Partner Datasets

MAP
CAARP
**Bonza**
AFIRC
Circus Oz
AusStage
AMHD
DAAO
AustLit
AWR
ADB
DoS
EOAS

**AUSTLANG**
Mura

**Media**: film, cinema, theatre, newspapers, magazines, advertising, music, live performances

**Biographical**: artists, designers, writers, significant people, scientists, Sydney demographics

**Indigenous languages**
Integrating Meaning

- Align and link the contents of the partner datasets using reference ontologies and vocabularies
  - The characteristics of the datasets determine the ontological components selected and the major entities
  - Major common entities identified as: people, organizations, events, relationships, places, dates, resources, and subjects
  - Reference ontologies being used or considered: CIDOC-CRM, FOAF, FRBR, FRBR-OO, BibFrame and PROV-O
- Express the combined data as Linked Data in RDF format
Integration Workflow

Original Data Sources

- DAAO
- ADB
- AWR
- EoAS
- Other DBs

HuNI RDF Ontology + Data Store

Linked Open Data Cloud

- FRBR-OO
- CIDOC-CRM
- SKOS
- ? Other ontologies that may be needed

- SPARQL Query
- User Interface
- RDF Tools

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
A researcher with a HuNI account will be able to:

- Search the HuNI data
- Save their search results as a private collection
- Refine their collection through additional searches
- Analyse and annotate their collection with their own assertions and commentary
- Export their collection for further analysis
- Publish and share their collections and analyses
Researchers will be able to:

- perform a “deep search” of the graphs in the RDF Triple Store;
- browse by high-level facets.

The large-scale aggregation of Linked Data makes explicit the relationships and connections between records across all the partner datasets, enabling the researcher to construct more complex semantic queries.
HuNI: Technical Overview

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.
Unlocking and uniting Australia’s cultural datasets

Introducing HuNi

The HuNi project is combining Australia’s most significant cultural datasets to build a national virtual laboratory. Using the lab, you will be able to:
- Conduct searches across more than two million authoritative records.
- Save your search results as your own collection and refine or expand your collection through additional searches.
- Analyse and annotate your collection.
- Export, publish and share your collection and analysis.

Find out more:
- What is HuNi?

News

Spreading HuNi

Recent blog posts

LODLAM Summit: building a web of knowledge
Questions aplenty!
Charting Australia’s media history

Upcoming events

User Interaction built on Library Linked Data (LALLD) Workshop
16 Aug 2013

HuNi 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