

Submission

2016 National Research Infrastructure Roadmap

Capability Issues Paper

Name	Professor Deb Verhoeven
Title/role	Director, HuNI
Organisation	Deakin University
Name	Dr Toby Burrows
Title/role	Product Owner, HuNI
Organisation	University of Western Australia

General Comments

This submission draws on our experience in developing the Humanities Networked Infrastructure ([HuNI](#)) as a national Virtual Laboratory under the Australian government's NeCTAR (National e-Research Collaboration Tools and Resources) program. HuNI is the largest humanities-specific research infrastructure capability in Australia, combining data from 33 of Australia's most significant cultural datasets. These datasets comprise more than 17 million authoritative records relating to the people, organisations, objects and events that make up Australia's rich cultural heritage. HuNI has been developed as a partnership between 16 public institutions, led by Deakin University and including major national institutions like AIATSIS and the National Library of Australia.

The functionality and data model deployed in HuNI are based on a needs analysis carried out across researchers from the many disciplinary groups contributing data. Users of HuNI are able to discover content, make connections across a wide variety of datasets, build their own datasets, and share these connections and datasets with other users. HuNI addresses research infrastructure issues of fundamental importance to the humanities (and of relevance to some STEM disciplines):

- How to support the nonlinear and recursive research methods practiced in the humanities;
- How to enable heterogeneous metadata to talk to each other without losing the unique nature of each disciplinary source;
- How to enable researchers to create their own classifications and categorisations;
- How to enable researchers to create their own links between data;
- How to offer suitable methods for addressing questions of gender, race and culture within new knowledge structures;
- How to re-define humanities "data" in an environment where a standard, "one-size-fits-all" approach is often taken to research data management;
- How to enable humanities researchers to create their own data and incorporate them into a larger aggregated pool of data;
- How organised information systems can encompass a humanities disposition for diversity, co-existence, complexity, interpretation and contestability.

HuNI enables the participation of users from the wider community, alongside the researchers who are the primary users of the Virtual Laboratory.

Question 1: Are there other capability areas that should be considered?

We are pleased to see that Understanding Cultures and Communities is now positioned in the list of capabilities.

Question 5: Should research workforce skills be considered a research infrastructure issue?

HuNI has adopted a “user centred” approach to information and data design that has had significant effect on addressing lower levels of data skills in that sector. For example, humanities data has in the past been developed by individual researchers or within highly specific disciplines. In this context, skills in data interoperability were not always a priority despite the evident gains.

Research infrastructure such as HuNI is by definition connective and scaled developments such as HuNI directly address these traditional humanities practices as a skills deficit. HuNI has been instrumental in producing solutions for humanities researchers across the research “pipeline” – providing opportunities to lift skills in data structuring, data curating, data analysis, data scholarship and data publication.

Question 6: How can national research infrastructure assist in training and skills development?

Our answer to this question relates to the short-term, development based funding in this sector (e.g. projects funded on an annual basis, dependent on LIEF funding). In this approach many experienced staff move on quickly; projects lose their skills base and the expertise and knowledge transfer their staff offer. This is inefficient and has led to the sense of constant fragmentation and re-invention in the field. Longer and more sustainable funding would have the benefit of ensuring a more durably skilled workforce in the humanities.

Question 7: What responsibility should research institutions have in supporting the development of infrastructure ready researchers and technical specialists?

Researchers need training in data management that is specific and appropriate to their discipline; institutions are potential sources of this training, in collaboration with research infrastructure providers.

Question 8: What principles should be applied for access to national research infrastructure, and are there situations when these should not apply?

Adequate and stable funding, including for national and state cultural collecting institutions, to ensure sustained access.

A National Research Infrastructure Roadmap needs to ensure that the products, including the content and community connections resulting from government funded initiatives, remain freely accessible and discoverable.

Clear recognition of the wider community, and non-institutional researchers, as end users of research infrastructure for Understanding Cultures and Community.

Question 13: In considering whole of life investment including decommissioning or defunding for national research infrastructure are there examples domestic or international that should be examined?

There have been numerous attempts over the past two decades in Australia to establish and run government funded national digital repositories, portals and communities of practice in education for the purpose of collaboration and sharing of educational research and resources, such as Education Network Australia (EdNA) and the Australian Flexible Learning Framework. Such services have provided a proven capability for delivering national educational research infrastructure over many years. However, once funding has been withdrawn from these services, they disappear and the content and communities of practice supported by these services are lost to future generations of educators and educational researchers.

Understanding Cultures and Communities

Question 24: Are the identified emerging directions and research infrastructure capabilities for Understanding Cultures and Communities right? Are there any missing or additional needed?

Our view of the relationship between cultural collecting institutions and research infrastructure recognises the crucial role that the collecting institutions have in providing digital materials and data that can then be consumed or reused by researchers in research infrastructure. This is a critical distinction. As the document acknowledges, the HASS infrastructure and datasets identified in the document illustrate the range of datasets forming research infrastructure capabilities for Understanding Cultures and Communities, but the list is far from exhaustive and does not include university-based research collections.

The relationship of the national collecting institutions to the national research infrastructure is an important one which deserves fuller and careful consideration. The issues paper as it is written assumes that collecting institutions are “vital research infrastructure”, but this is not necessarily how they function for researchers, who are only one of their user groups. The national institutions are information providers with users from across the entire community, and their funding and services are directed towards much broader goals. How their collections can be drawn into a “national digital humanities capability” for research requires careful thinking and consultation with researchers themselves.

We would support a “national digital humanities capability” which focuses on reuse, integration, interoperability and data curation, as well as providing a platform for building tools and techniques designed to address research problems. We believe research infrastructure in the humanities needs to do more than provide better access to digitised cultural collections and making them more discoverable. Two fundamental components of this humanities research infrastructure should include:

- Open Data capability: collecting institutions (and other data providers) should expose their metadata and digital objects in formats which can be reused for research infrastructure services (using appropriate AusGOAL licences). This would include the IIIF (International Image Interoperability Framework) for images; metadata in formats which can be consumed by Linked Open Data services; and text in formats which are suitable for text-mining and text analysis (the HathiTrust in the U.S. is an important example of this);
- Knowledge Graph capability: vocabulary and ontology services, Linked Open Data services, visualisation services, etc. – for linking and integration of metadata from disparate sources (both cultural institutions and specialist databases).

Both of these are foundational components of the HuNI Virtual Laboratory’s development.

Although we broadly support the document’s statement that “Research infrastructure-like activities currently undertaken at national cultural institutions need to be supported and recognised as core national infrastructure, as important as any other research infrastructure holding, and just as irreplaceable”, we also note that the future development of UCC infrastructure is an opportunity for the integration of collecting institutions’ repositories with researcher-led infrastructure such as HuNI, and its and other Virtual Laboratories’ development of digital research tools, techniques and methodologies.

The [CLARIAH](#) (Common Lab Research Infrastructure for the Arts and Humanities) facility in the Netherlands, which is being funded for five years (2015-2019) under the NWO’s National Roadmap for Large-Scale Research Facilities, provides a relevant and useful model for the future. CLARIAH is combining textual, audio, visual and structured data into a “Common Lab” environment where researchers can deploy a variety of tools. Semantic interoperability through Linked Open Data is a crucial component of the platform. In an Australian context, this would involve a combination of the capabilities of the HuNI and Alveo Virtual Laboratories, with elements of the Australian Data Archive, ATSIDA and PARADISEC.

Question 25: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

The emergence of networked and linked data structures in the humanities (in which HuNI is a leading proponent) suggests many future collaborations particularly across the linked data ecology.

HuNI has been contacted by many European partners (full realisation of these collaborations is currently restricted due to the limited project-based nature of HuNI funding). These include:

HistoGraph: <http://histograph.eu/>

Europeana: <http://www.europeana.eu/portal/en>

DBPedia: <http://wiki.dbpedia.org/>

HuNI's approach to vernacular (unstructured) data ecologies is also being used as a model by the commercial publisher Elsevier with many potential applications.

Question 26: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Understanding Cultures and Communities capability area?

It is important to note the document's recognition of digitisation as Underpinning Research infrastructure, a case that researchers in the sector have been arguing for well over a decade. Having secured this recognition however, more attention needs to be paid to enhancing and systematising existing digitisation of cultural artefacts (from media to historical material) and how these digital assets can be re-used and analysed in order to produce innovative research solutions in the humanities.

Underpinning Research Infrastructure

Question 30: Are the identified emerging directions and research infrastructure capabilities for Underpinning Research Infrastructure right? Are there any missing or additional needed?

As noted above, it is pleasing to note the document's acknowledgement of digitisation as infrastructure, and its proposal that a digitisation initiative would benefit from national coordination and funding with a research focus. Such a development would be a critical contribution to the continuing development of digital humanities research effort.