The Tropical Data Hub (TDH) – A virtual research environment for tropical science knowledge innovation and discovery

Trina Myers[1], Jarrod Trevathan[1], Ian Atkinson[1], Rob Cook[2], Jeremy VanDerWal[1].
[1]James Cook University, Townsville, Australia
{Trina.Myers, Jarrod.Trevathan; Ian.Atkinson, Jeremy VanDerWal}@jcu.edu.au.
[2]Queensland Cyber-Infrastructure Foundation (QCIF), Brisbane, Australia
Rob.Cook@qcif.edu.au

The *Tropical Data Hub* (TDH) is a platform to serve data sets related to tropical research from a single virtual location that will enable researchers, managers and decision-makers to collaborate around the data. Currently focusing on the tropical sciences, knowledge, and innovation spanning the top-end of Australia, the TDH will, over time, be expanded to cover a greater area of the world's tropics.

Up to 80% of animal and plant species and 92% of world's coral reefs are located in the world's tropics. The challenges of population growth, urbanisation, climate change, and biodiversity loss in these areas can only be addressed successfully with cross-disciplinary approaches [1]. The TDH philosophy is to span traditional 'vertical' research disciplines and enable 'horizontal' research. Specifically, vertical research is the traditional discipline and data-specific research paradigms are conceptual silos of concentrated research efforts. In contrast, horizontal research spans a cross-connect through disciplines, research methods, data resources and experimental techniques to enable synthesis of a diverse range of disciplines and data. Here the scope of an investigation or line of enquiry could span across scientific and economic approaches. For example, questions posed by the scientist may involve complex component models for integrated hypotheses and/or assessment and understanding of how increases and decreases affect ecosystems. In contrast, policy makers can receive clearer pictures of the sensitivity of ecological impacts from population growth from these models.

The internally and externally stored data will be published via the TDH in a form that is conducive to the linked data initiative. The linked data movement defines the processes and best practises for publishing and connecting structured and semi-structured data on the Web [2]. The TDH is intended to complement existing data repositories and stores, and be a banner under which research interests in the tropical world can be concentrated. The hub will provide a data hosting infrastructure to host significant national and international data sets, but significantly also be portal to other primary data sources. Notably, the portal's added key functionality will be the amalgamation of disparate and unrelated data exposed for the purpose of harvesting metadata and facilities to search across all data sources. Interested parties such as Australian state, commonwealth and university research programs and/or management groups will then benefit from the synthesis of linked cross-discipline data, information and knowledge available via the TDH.

The TDH architecture is being developed in stages due to the extensive scope (Figure 1). Currently, the development is focused on four major projects in varying stages of completion: the creation of the main portal; the metadata and data storage infrastructure; data and metadata discovery, capture, collection, curation and publication capabilities; and data analysis tools. The TDH portal project has created an aggregator to provide pointers to collections and tools to register, identify, curate, deposit and access tropical data. There are two complimentary projects to support data and metadata advertisement and discovery. The first aims to "seed the commons" by identifying and creating collection descriptions for a wide range of existing datasets for availability in *Research Data Australia* via the *Australian National Data Service* (ANDS) [3]. The second focuses on data capture and description software to automate the capture data and metadata from new and existing data sources. For example, real-time sensor data from next generation wireless sensor network currently being deployed in both terrestrial and marine environments such and the Daintree Rainforest Observatory. The datasets are to be published through searchable data/metadata repositories to ensure users have maximum exposure to the data. The TDH data analysis tools will expose data for internal and external automated linkage and integration, correlation, querying, and visualisation across different domains of research. These *Virtual Research Environment* services and tools are to serve data sets from a single virtual location to enable researchers, managers and decision-makers to collaborate around the data.

Currently, two exemplar demonstration sites have been produced to exhibit data integration for data visualisation and assessment of land use and species distribution. The first is a portal for the evaluation of significant land use and species distribution data in Northern Australia. In this use-case, geo-hydrological data (vector data) and bird species distribution (raster data) in the wet tropics of North Queensland is incorporated [4]. The *Geographic Information System* enables search facilities to overlay primary hydrological data on river attributes with the migratory movement and refugia of hundreds of bird species over a period of time. The information derived can be used in the management and conservation decision-making of stream ecology natural resources.

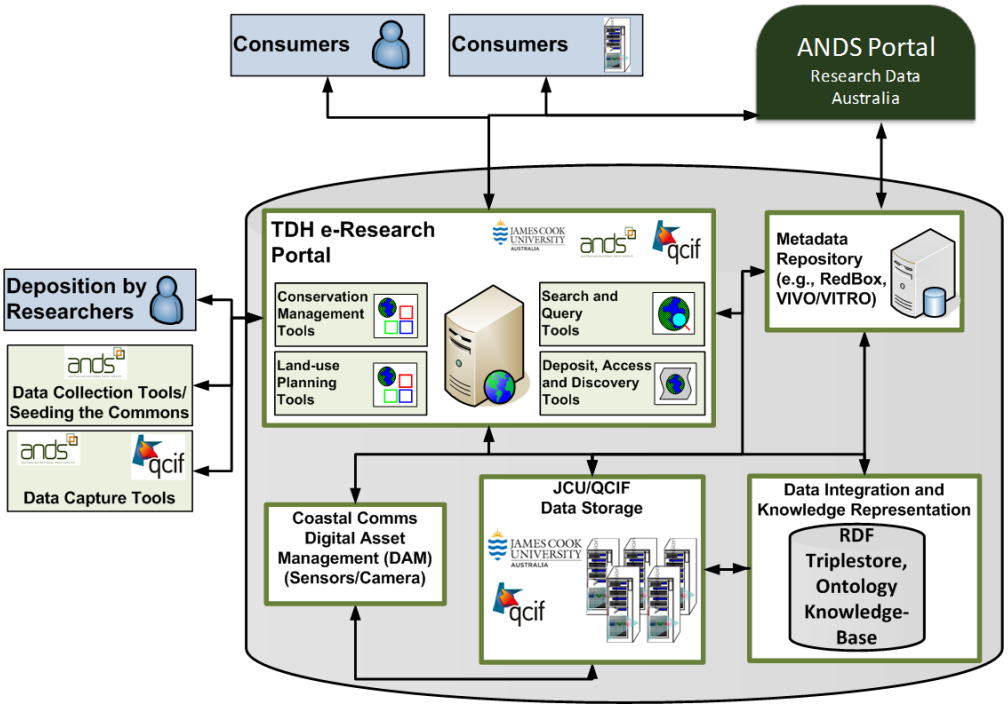

Figure 1: The TDH architecture

The second exemplar is the *Wallace Initiative* (WI) demonstration site which models the potential impacts of climate change. The WI is a global effort to rapidly assess the projected magnitude of change and the impacts on global biodiversity [5]. It brings together experts on climate change (Tyndall Climate Change Centre, UK) and bioclimatic modelling (TDH, Australia and the Centre for Tropical Agriculture, Colombia), with global datasets on biodiversity (GBIF, Denmark) and high performance computing facilities (*James Cook University and the Queensland Cyber Infrastructure Foundation*). The current model shows the impact on approximately 50,000 terrestrial plant and animal species. The models are used to calculate projected changes in percent species richness. The focal areas are likely to be refugia under various degrees of temperature change, and areas likely to undergo the greatest loss. This information can provide guidance to the *World Wildlife Foundation* and its partners on how they may need to adapt to avoid biodiversity loss. The web portal allows other practitioners, decision makers, and eventually, the public, to have access to these climate change projections.

The future development of research in Tropical science will be an important tool in the discovery of phenomenon such as climate change impacts, the study of endangered or noxious flora and fauna, decision-making for land use and marine management. The TDH is proposed as an 'open' portal, with contributors submitting data sets and other content in an open and collaborative way. Over time it is expected that the Hub will incorporate elements that go beyond research data, and have a role in public dissemination of research to government, industry and the public. Our vision is a Tropical Hub that will be the centre of a range of eResearch services and outputs from organisations such as the *Australian National Data Service* (ANDS) [3], the *Terrestrial Ecosystem Research Network* (TERN) and universities involved in Tropical research and development.

REFERENCES

1 IPCC, "Climate change 2007: the physical basis," Cambridge Univ. Press, Cambridge, U. K. 2007.

2 C. Bizer, T. Heath, and T. Berners-Lee, *Linked data-the story so far*, Int. J. Semantic Web Inf. Syst., 2009 **5**(3): p. 1-22.

3 *The Australian National Data Service.* http://ands.org.au/ accessed May 2010.

4 S. Januchowski, J. E. VanDerWal, B. Pressey, and A. Edwards, *Characterizing errors in digital elevation models and estimating the financial costs of accuracy*, International Journal of Geographical Information Science, 2010 **24**(9): p. 1327-1347.

5 J. T. Price, R. F. Warren, J. Vanderwal, L. Shoo, J. Ramirez, A. Jarvis, and S. Goswami, *Making Scientific Data Available to Adaptation Practitioners-the Wallace Initiative*, in *American Geophysical Union, Fall Meeting 2010*, - 1 1. 2010,