Worldwide Universities Network (WUN) Web Observatory

Applying Lessons from the Web to Transform the Research Data Ecosystem

Simon Price
University of Bristol, UK
simon.price@bristol.ac.uk

Wendy Hall, Graeme Earl,
Thanassis Tiropanis,
Ramine Tinati, Xin Wang,
Eleonora Gandolfi
University of Southampton, UK
wh, tt2, rt506@ecs.soton.ac.uk

Jane Gatewood
University of Rochester, USA
jane.gatewood@rochester.edu

Richard Boateng
University of Ghana, Ghana
richboateng@ug.edu.gh

David Denemark
The University of Western
Australia, Australia
david.denemark@uwa.edu.au

Alexander Groflin
University of Basel, Switzerland
alexander.groeflin@unibas.ch

Brian Loader
University of York, UK
brian.loader@york.ac.uk

Maxine Schmidt,
Marilyn Billings
University of Massachusetts
Amherst, USA
mschmidt@library.umass.edu

Gerasimos Spanakis
Universiteit Maastricht,
The Netherlands
jerry.spanakis@maastrichtuniversity.nl

Hussein Suleman
University of Cape Town,
South Africa
hussein@cs.uct.ac.za

Kelvin Tsoi
The Chinese University of
Hong Kong, China
kelvintsoi@cuhk.edu.hk

Bridgette Wessels
Newcastle University, UK
bridgette.wessels@ncl.ac.uk

Jie Xu, Mark Birkin
University of Leeds, UK
j.xu, m.h.birkin@leeds.ac.uk

ABSTRACT

The ongoing growth in research data publication supports global intra-disciplinary and inter-disciplinary research collaboration but the current generation of archive-centric research data repositories do not address some of the key practical obstacles to research data sharing and re-use, specifically: discovering relevant data on a global scale is time-consuming; sharing 'live' and streaming data is non-trivial; managing secure access to sensitive data is overly complicated; and, researchers are not guaranteed attribution for re-use of their own research data. These issues are keenly felt in an international network like the Worldwide Universities Network (WUN) as it seeks to address major global challenges. In this paper we outline the WUN Web Observatory project's plan to overcome these obstacles and, given that these obstacles are not unique to WUN, we also propose an ambitious, longer-term route to their solution at Web-scale by applying lessons from the Web itself.

Keywords

Research Data Management, Data Science, Social Machines.

1. INTRODUCTION

The Worldwide Universities Network (WUN) is the most active global higher education and research network with 90 active research initiatives, engaging over 2,000 researchers and students collaborating on a diverse range of projects. These initiatives are committed to addressing some of the world's most urgent challenges and are supported by prolific partners such as the United Nations Foundation, World Bank, OECD and World Health Organization. WUN research focuses on four themes that address the following globally significant challenges.

- **Responding to Climate Change.**
  Explores sustainable approaches to food and environment security, addressing the scientific, cultural, health and social issues that govern our response to a changing climate.

- **Public Health (Non-Communicable Disease).**
  Emphasizes a life-course approach to addressing non-communicable diseases (NCDs), especially but not exclusively, in low and middle income countries and transitioning populations.

- **Global Higher Education and Research.**
  Proposes policy reforms to address the sources, mech-

anisms, and social structures that give rise to access, mobility and investment challenges for international research and education.

- **Understanding Cultures.**
  Facilitates the understanding of some of the principal consequences of globalisation for cultural identities, migration, population, trans-disciplinary indigenous research, global digital cultures, and heritage.

Collaborative and multi-disciplinary research programmes are a crucial component of WUN's response to these global challenges, which depends on experts from across the globe working together, sharing both open and sensitive data between and across disciplines.

The sharing and citation of research data that support academic publications is increasingly an expectation of journals, funders and researchers. The proliferation of institutional, national and discipline-specific research data repositories provides researchers with a wide choice of venues in which to publish their research data - so much so that finding and accessing relevant research data without a direct link can be difficult. The Australian National Data Service runs a national discovery service for Australian research data, and the UK's JISC is establishing a similar service for UK research data. So, while there are now a multitude of repositories publishing metadata descriptions of, and links to, their research data, the global discovery and re-use of data requires researchers to search multiple metadata directories within specific territories or disciplines.

This situation echoes the early days of the Web when it was still feasible for a small number of "centralised" organisations to curate usable directories describing all the websites in an institution, a nation, a discipline or the world.

2. WUN WEB OBSERVATORY

The Web Observatory infrastructure [3, 4], developed under the auspices of the Web Science Trust, brings the principles of the early Web to this emerging research data ecosystem, with the same goal of transformational growth through decentralisation [1]. Below we describe three components to this infrastructure that the WUN Web Observatory project seeks to exploit.

2.1 Virtual Research Data Repositories

Anyone can establish a Web Observatory to create their own *virtual research data repository* with its own, possibly highly focused, scope and self-determined rules for access and sharing. At a foundational level this enables researchers and organisations to create individualised or project-specific research data directories and to maintain control over the description, ownership and access to their own research metadata. Indeed, one of the objectives of the WUN Web Observatory project is to create a WUN research data directory by establishing a network of Web Observatories cataloguing the institutional research data repository and other major holdings of all the WUN member universities. The resultant self-updating and readily searchable directory of WUN research data will facilitate research data sharing within and between WUN universities and support interdisciplinary and inter-institutional research.

2.2 Management of 'Live' Research Data

However, the greater significance of creating the *WUN virtual research data repository* is that its constituent network of Web Observatories provides the requisite critical mass for the realisation of the WUN Web Observatory project's transformational vision: to position the WUN as global leaders in the exchange and re-use of research data between disciplines and institutions, driving forward research underpinning WUN's priority Global Challenges and Cross-Cutting Themes while, at the same time, bootstrapping the worldwide development and adoption of the Web Observatory infrastructure envisioned and supported by the Web Science Trust. Establishing a network of Web Observatories facilitates the bottom-up adoption of Web Observatory infrastructure by WUN researchers for the creation and widespread use of virtual research data repositories as the default means of research data discovery and exchange. Crucial to its appeal, and unlike the current generation of research data management software, Web Observatory infrastructure supports the management of active, 'live', research data, not just archival post-publication research data.

2.3 Secure Access Control and Attribution

Using a technical trick, known as "reverse proxying", Web Observatories enable authentication and access control to decentralised research data - regardless of where that data is held, whether it is in an archival repository or in working, active research stores, digital asset management systems or in arbitrary cloud-hosted platforms or even in streamed real-time data feeds. This distinguishes a virtual research data repository from a research data directory or a traditional research data repository and is the core innovation of the Web Observatory concept that the WUN Web Observatory project seeks to investigate and develop through a range of research demonstrator activities, described below.

3. RESEARCH DEMONSTRATORS

To explore research questions across the four Global Challenges, the WUN Web Observatory project established four corresponding demonstrator activities that also support the Cross-Cutting Themes as depicted by numbered circles and themes in the annotated WUN research matrix of *Figure 1*.

Each of the four research demonstrator activities explores different aspects and configurations of the Web Observatory infrastructure to enable and support WUN research as outlined below.

3.1 Demonstrator 1: Online Learning

Aggregates datasets from different online learning platforms across the WUN and provides analytics on those datasets, quantifying the diverse usage patterns of Massive Open Online Courses (MOOCs) to reveal success criteria for online learning and to provide insights into future directions[1].

3.2 Demonstrator 2: Ageing and Well-being

Combines data from established sources, such as cancer registries, with newer (often personalised) sources, such as wearable sensors, mobile devices and crowdsourcing, to enable visualisation and analysis of ageing patterns and poor

---

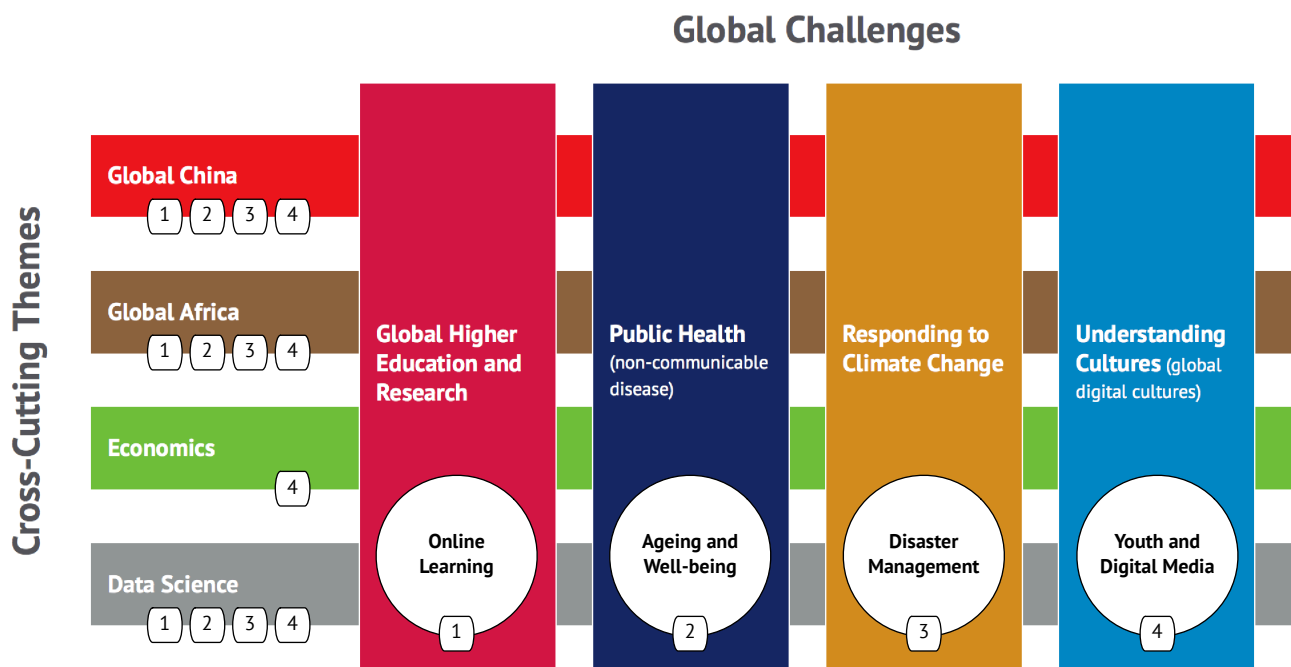[1]Online Learning lead: Spanakis, Universiteit Maastricht.

Figure 1: Mapping to WUN Web Observatory demonstrators to Global Challenges and Cross-Cutting Themes

health, disease trend projection, and perceptions of ageing well[2].

3.3 Demonstrator 3: Disaster Management

Enables and investigates social and technical mechanisms for Citizen-Driven Disaster Management, where individuals provide their own data in a decentralised system to proactively support emergency response and planning prior, during, and post natural and urban disasters, while retaining control and privacy of their own data[3].

3.4 Demonstrator 4: Youth and Digital Media

Links and analyses disparate WUN datasets to explore the patterns and impact of youth reliance on social media and digital information platforms for their political involvement, and the consequences for young citizens' attitudes toward government, citizenship, politics and civic engagement[4].

4. SUMMARY

Building on experience from these four research demonstrators, the WUN Web Observatory project[5] aims to instantiate a collection of virtual research data repositories, enable effective management of 'live' research data, and control of its access and attribution; thus becoming the default pathway for WUN's researchers working in pursuit of its Global Challenges. Such widespread international adoption

---

[2]Ageing and Well-being lead: Tsoi, The Chinese University of Hong Kong.
[3]Disaster Management lead: Tinati, University of Southampton.
[4]Youth and Digital Media lead: Denmark, The University of Western Australia.
[5]Project website – `http://wun.ac.uk/wun/research/view/web-observatory-project`

significantly advances the project's longer-term ambition to bootstrap the worldwide development and adoption of Web Observatory infrastructure to the benefit of all research [2].

5. REFERENCES

[1] W. Hall and T. Tiropanis. Web evolution and web science. *Computer Networks*, 56(18):3859–3865, 2012.
[2] C. Phethean, E. Simperl, T. Tiropanis, R. Tinati, and W. Hall. The role of data science in web science. *IEEE Intelligent Systems*, 31(3):102–107, May 2016.
[3] T. Tiropanis, W. Hall, J. Hendler, and C. de Larrinaga. The web observatory: a middle layer for broad data. *Big Data*, 2(3):129–133, 2014.
[4] T. Tiropanis, W. Hall, N. Shadbolt, D. De Roure, N. Contractor, and J. Hendler. The web science observatory. *IEEE Intelligent Systems*, 28(2):100–104, 2013.