Professor Paul Arthur

Chair in Digital Humanities and Social Sciences, Edith Cowan University, Australia
paularthur.com / @pwlarthur / paul.arthur@ecu.edu.au

DADH2019 conference
Taipei, December 2019

How Large-Scale Data Aggregation is Transforming Society and Enabling New Research Practices

[SLIDE: title slide]

I am very grateful and honoured to have been invited to speak at the tenth international conference of Digital Archives and Digital Humanities. I would like to thank the conference organisers, and the National Taiwan Normal University.

My presentation is in two parts. In the first part I give an overview of the impact of digital transformations in society on a global scale. We are in the midst of a data revolution that is influencing the daily life of much of the world's population so rapidly and profoundly that it is impossible to grasp the extent of its effects. While instant access to proliferating stores of information is helping us better understand our world, it is also shaping us as human beings. Aggregated data is having a growing impact in practically every area of our lives. I will explore how the digital environment is fundamentally changing our identities and social behaviours.

In the second part of the presentation I discuss how digital humanities is benefitting from the very same principle of increased data aggregation and linkage that we see in society generally, but applied to research. Some of the largest and longest running digital projects have built up their content and collections over more than two decades, aggregating and enriching vast amounts of digitised material. The success of these projects can be measured in terms of their growing scale and coverage, and yet most have remained as stand-alone projects. More recently there has been a global trend towards developing research infrastructures that provide frameworks for connecting together digital projects and collections by enabling data interoperability. I will refer to some of the most important Australian digital humanities projects that have contributed to the development of the field nationally. I will also discuss examples of Australian digital research infrastructures that seek to aggregate and integrate multiple resources and archives to produce new knowledge and understandings of culture and history.

Part 1

[SLIDE: world map network image]

Over recent decades innovations in information technology and computing have been proliferating with breath-taking speed. Global digital connectivity has triggered a transformation whose impact is greater than that of any other innovation in the history of technologies of communication. Digital systems and communication networks have dramatically reshaped the way we relate to each other, how we perceive ourselves, what we value, and the ways we interact with the world around us. Even those without access to the Internet are affected in that the world they inhabit is being increasingly controlled, regulated and monitored using digital technologies. There is little of our world that remains unaffected.

These impacts are demonstrated by many examples including: satellite navigation systems; transport and logistics infrastructure; financial markets and banking systems including new digital cryptocurrencies; digitally-based user-driven business models; online political campaigns; e-government and open data; medical robots and health data records; environmental sensing, and outer space exploration. Even wars are enacted via the Internet through cyber-attacks, automatic defence networks, drone strikes and digital surveillance. And now many people live with early forms of machine learning and artificial intelligence in smart homes, and via smart devices.

[SLIDE: medical polymer image]

Huge medical advances have also taken place especially during the past decade, driven by digital technologies, such as the implanting of microcomputer chips under the skin and the production of prosthetics via 3D printing. Readily available now in the mass market are body metrics textiles and other wearable electronics that react to changes in physiology. New technologies are plugging us into a non-human realm of constant interactivity between ourselves and digital devices and monitors and blurring the line between human and non-human. Many of these 'things' are not only around us, they are within us – in our bodies, under our skins and in our sensory systems. The Human Cell Atlas project is aiming 'to map the human body's 35 trillion cells… to decipher the types and properties of every cell, the project will attempt to work out exactly what we are made from'. In what may be thought of as a sinister twist, Google's parent company Alphabet is encouraging people to donate their DNA.

These examples have in common the capacity to perform ever more complex functions, at ever higher levels of speed and efficiency, and with less and less direct human hands-on involvement. And underlying these developments has been a massive increase in the amount of data being generated and collected every day across the world. On an average day, by some estimates, 'human beings generate 2.5 million trillion bytes of data'.

Most of the advanced applications of digital technologies I have just referred to rely on the aggregation of 'big data' to model and understand aspects of human behaviour and to inform the design of new systems. However, actually reaching this vast scale of mass data collection and aggregation is a very recent phenomenon.

[SLIDE: image of head]

It wasn't until the first smart devices and the release of the iPhone around 2007 that there was a significant critical mass of Internet-connected users across the world. Now we are more than ten years on from this time. Technology is part of so many people's lives. Even though there continue to be huge numbers of people without access to the Internet and digital technologies, statistics show that there has been an exponential growth of Internet connectivity, which will continue into the future.

There is no doubt that we are in the midst of a data revolution that has penetrated the daily life of much of the world's population so suddenly and deeply that it is impossible to grasp the full extent of its impact on the concepts of self and identity in the digital world.

The field of digital humanities has recently begun to address this most fundamental of human questions: What does it mean to be human in the digital age? How are identities and behaviours changing, and what kind of future does this point towards? These questions are very much in line with this conference's inclusion of the theme of 'Digital Humanities and Social Trends'.

In this first part of the talk, with these questions in mind, I look back over the past few remarkable decades of digital innovation to the present day.

When the term 'digital humanities' was coined around 2004 in the title of the well-known edited collection *A Companion to Digital Humanities*, it marked a major shift in the field that had previously been referred to as 'humanities computing'. The shift was towards a wider understanding of how digital technology and media was influencing not just research practices, but everyday lives.

Mass connectivity has, for many people, removed the traditional restrictions to communication that were posed by geography. All kinds of borders, physical and conceptual, have been broken down by the Internet.

[SLIDE: Babbage analytical engine image]

The history of the development of computers goes back to the first experiments by the likes of Charles Babbage, an English mechanical engineer who pioneered the concept of a programmable computer in the early 19th century. As in the case of most computers prior to the Internet era, his analytical engine was designed for performing calculations, not for communication or information sharing. Very few people used or even saw such machines.

For many people in developed countries (I speak here from the western perspective I am familiar with), their first encounter with computers was in the 1980s when desktop computers entered the mainstream in businesses and even the home. These were not yet communication devices, although they were starting to be linked on localised networks.

[SLIDE: http image]

During the same period, something was quietly happening that would start the digital revolution. The network we now know as the Internet was being established. The Internet was initially conceived of for military and research purposes, and at first only connected a limited number of sites in the United States; but it wasn't long before the hypertext protocol and the World Wide Web

were established and new nodes worldwide opened the system to users across the globe, albeit still only a privileged few. If we compare the late 20th century, when the first personal computers and the early Web appeared, with the second decade of the twenty-first century, one of the greatest differences is that the specialist computing power that was only available to a very small number of people, almost entirely in the domain of the sciences, is now within easy reach of many researchers and even the general public.

[SLIDE: device image]

The supercomputers of the last century have been progressively compressed and miniaturised to such an extent that their equivalents now fit into devices that we can hold in the palm of the hand. It was in the 1990s that the first laptops became readily available. By the turn of the century, there was much discussion of technological convergence of telephony, computing and traditional media channels such as TV and radio. We are living in a time of truly ubiquitous computing. Our devices are designed to be always-on, and so too is society, constantly connected.

[SLIDE: software apps image]

Another major change is that sophisticated software and services that didn't yet exist or were extraordinarily expensive are now often available free of charge. At the beginning of the 21st-century there were approximately 350 million people connected to the Internet across the world, but by the end of the first decade this number had increased to more than 2 billion; by end of 2017 it was over 4 billion at last count. During the same period mobile phone subscriptions skyrocketed from 750 million to over 6 billion worldwide. According to a 2016 Ericsson Mobility Report, over 90% of the world's population will be covered by mobile broadband networks by 2021. If this momentum is maintained most of the earth's projected population of 8 billion people will be online by 2025.

[SLIDE: on button image]

It is not only humans who are caught up in the global web of digital connectivity but also an exponentially expanding number of "intelligent" devices in what is known as the Internet of Things. There will be 30 billion connected intelligent devices by 2020. Last year, for the first time, the number of smart devices was greater than the world's population. It is astonishing how quickly this world transforming phenomenon has been accepted as normal and routine. Each day digital technologies take a firmer hold on the basic systems that underlie the functioning of modern global societies.

[SLIDE: Web 2.0 image]

In the early years of the 21st century, from around 2004, social media entirely reconfigured human interactions on the global scale. Web 1.0, as it is often retrospectively referred to, had created new communities, but Web 2.0 took this further to create, as Van Dijck puts it in her classic book *The Culture of Connectivity* (2009), 'interactive, two-way vehicles for networked sociality'.

[SLIDE: crowdsourcing image]

Crowdsourcing – a term coined in 2006 by *Wired* writer Jeff Howe – harnesses the collective power of multiple users or actors in a system or network. The same power, of the crowd, enables

crowdfunding, and other forms of what we may call collective intelligence, or connected action. The same collective energy can also be directed at social causes. With a powerful enough message, marketers, activists, politicians, and lobbyists can summon vast audiences in a matter of minutes, swaying mass public opinion. The breaking of news by the likes of Wikileaks even has the power to affect global financial markets.

The collective power of social media to mobilise social action was demonstrated memorably in the Arab Spring in late 2010, and other mass activism since, giving collectives and communities voice that was not previously possible. The recent #metoo movement has shown the power of ground-up social media to expose and bring down celebrities who once dominated the one-way communication channels of traditional broadcast media.

[SLIDE: surveillance binoculars]

At the same time as accessing this vast realm of accumulating and circulating data via our networked devices, we are also contributing to it with every click or touch – adding bits and pieces of ourselves. As millions of users logged into smart phones, vast new data sets began to be generated continually, producing 'big data' on the movement and behaviour of people on an unprecedented scale and forming vast archives for potential data mining and matching. This opened the possibility for sophisticated mapping of data points and transactions simultaneously on the micro-macro scale. Digital data, it has been suggested, 'may be our era's microscope or telescope'.

As we traverse the Web, we leave a digital trail – and through this trail, whether we like it or not, we create online identities – and have them fashioned for us. Daniel J. Solove put it well in his 2004 book *The Digital Person*, in which he writes about the 'aggregation effect' in these terms:

> The digital revolution has enabled information to be easily amassed and combined. Even information that is superficial or incomplete can be quite useful in obtaining more data about individuals. Information breeds information…. Similar to a Seurat painting, where a multitude of dots juxtaposed together form a picture, bits of information when aggregated paint a portrait of a person.

[SLIDE: modern Seurat image]

Questions about the nature of the digital identities being generated and who has access to them and for what purposes are being asked with increasing urgency. The answers are not clear. Facebook, for example, has recently come under the spotlight for the fact that personal data from as many as 87 million user profiles was able to be accessed by their consulting company Cambridge Analytica, with huge political consequences.

The fact is that the more time we spend online, the more data that is gathered about our preferences based on search patterns; and in the process our online experiences become increasingly personalised and tailored to us by services such as Google provides. But this can also have negative effects, creating what are referred to as filter bubbles or echo chambers.

Running through both parts of my presentation is the theme of the increasing power of aggregated data. The examples in the first section of this talk show the power of aggregated data in the everyday world around us. As computing and digital media technologies have begun to influence all areas of life, their impacts have become the focus in many disciplines, especially in the humanities.

As I have indicated, this is now becoming an area of study in digital humanities specifically, with papers and panels at major conferences focused on questions relating to digital identity, human interactions with artificial intelligence, and data privacy, surveillance and ethics, to name only a few of the key areas.

Part 2

[SLIDE: e-humanities quotation]

In the next section of my talk I will discuss how the power of aggregated data is also enabling advanced research in the digital humanities field.

I will discuss the development of digital humanities in Australia, with reference to major projects that are national exemplars, as well as significant activities and initiatives that have formed a basis for the Australian field. In all these cases the projects began with modest goals. But the examples I will give have in many cases now been active and supported for two decades, and so they have reached a scale where they have become invaluable national research resources. Also linking the examples is a common theme: they all have the goal of capturing, preserving, building upon and articulating the richness of Australian culture and history for current and future generations. Their power is in the way they bring together data from multiple sources into large-scale project frameworks for public and research use. Most include a substantial digitisation component, or else draw on national digitised collections. They aggregate and enrich this data and in so doing provide vital online resources for humanities research in Australia.

A range of large-scale digital projects and events provided the context and impetus for the development of the digital humanities field nationally. As is the case in many parts of the world, such institutional activities tend to be associated with particular research centres or groups, and are typically collaborative enterprises. Major projects have helped to determine priorities and interests for the field as a whole and to foster communities of practice. In historical and disciplinary terms, many can be linked directly back to the humanities computing tradition that predated and provided the foundation for digital humanities as it is now known. The term ‘e-humanities’ was also sometimes used in Australia, as in the case of the slide being shown currently.

The launch in 2002 of the Australian e-Humanities Network and Gateway was a key milestone in the development of the field. This online database of digital projects in the humanities was the culmination of an Australian Research Council Learned Academies Special Projects scheme grant led by the Australian Academy of the Humanities, hosted by the University of Sydney with the University of Newcastle as a partner. Although it has since become inactive, in its time the portal gave exposure and recognition to a vast variety of digital projects that were pioneering examples of digitisation and of scholarly efforts to create enhanced digital resources for preservation and access in the early Web era. The original website described the field of e-humanities in these terms:

> As digital technologies rapidly transform the practice of humanities research, new arenas of theoretical inquiry engaged with this revolution are also emerging. Together these advances are coming to be known as the ‘e-humanities’, and in this field Australian humanities scholars are world leaders. (e-Humanities Gateway)

[SLIDE: Centre for Literary and Linguistic Computing]

Ground-breaking research in language and textual studies, for which Australia is internationally regarded, has been undertaken at the Centre for Literary and Linguistic Computing, established in 1989 at the University of Newcastle, where John Burrows pioneered early techniques in stylometry (the statistical analysis of literary stylistics). Burrows received the Busa Award for Outstanding Contributions to the field of Humanities Computing in 2001. The centre's continuing work has been led by Hugh Craig from 2001 to 2018. This foundational program has included development of techniques for stylistic analysis and authorship attribution including the Delta, Iota and Zeta methods, applied to Shakespeare and to a wide range of other textual sources and corpora.

[SLIDE: TimeMap]

The projects of the Archaeological Computing Laboratory (later Arts eResearch) at the University of Sydney developed by Ian Johnson, including 'TimeMap' from 1995 and those built on the Heurist reference database system since 2005, have featured an emphasis on enabling end users to create their own data management solutions and on web-based mapping with a temporal dimension.

[SLIDE: Digital Harlem]

They include the award-winning 'Digital Harlem', developed by four University of Sydney historians 'to visualize and explore the spatial dimensions of everyday life in Harlem during its heyday, 1915–1930'.

[SLIDE: Dictionary of Sydney]

They also include the 'Dictionary of Sydney', a project conceived in 2004 and launched as a born-digital web encyclopaedia in 2009, to provide a window into the history of Sydney.

[SLIDE: Expert Nation]

Most recently, this group has produced 'Expert Nation: Universities, War and 1920s & 1930s Australia'. Because the projects share the same Heurist reference database system, they effectively aggregate a very wide variety of cultural and historical materials.

[SLIDE: Austlit]

'AustLit' was started in 1999 by a consortium of universities, led by the University of Queensland and the University of New South Wales, initially with the National Library of Australia. Its stated purpose was to '1. to house a range of existing research and bibliographical projects relating to Australia's literary and print-culture history; and 2. to build the technical and intellectual infrastructure necessary to support further scholarship in the field'. The AustLit consortium, representing multiple universities and under the directorship of Kerry Kilner since 2002, has built upon this base to become what is now considered the world's most comprehensive record of a nation's creative writing across all forms and genres, and critical works associated with that output. AustLit has also developed a specialisation in Aboriginal and Torres Strait Islander writers through the BlackWords project and continues to support a diverse range of research, digitisation, and publishing related to Australian culture and storytelling.

[SLIDE: Austage]

Further projects initiated at the turn of the century included ‘AusStage’, the Australia Live Performance Database, started by the Australasian Drama Studies Association (ADSA) in 1999 to address ‘the need for research information in Australian theatre, drama and performance studies by building an index of performing arts events in Australia, and a directory of research resources on the performing arts’. The first phase of the project involved theatre scholars from eight Australian universities, with the Australia Council and industry partner Playbox Theatre. Through successive grants, AusStage has evolved over a 20-year period, with a recent focus on linking to international collections and visualising Australian live performance venues. The project is currently led at Flinders University by Julian Meyrick and managed by Jenny Fewster.

[SLIDE: South Seas Project]

The Consortium for Research and Information Outreach (CRIO, 2001–2009) at the Australian National University (ANU) was associated closely with the projects of Howard Morphy. With strengths in digital anthropology and filmmaking, this formed the foundation for the Digital Humanities Hub (now Centre for Digital Humanities Research), established in 2010. The centre is a leading research concentration and is indicative of the growing institutionalisation of the field nationally. CRIO was preceded by the development of web-based scholarly editing standards for history and heritage studies undertaken by the South Seas Project led by Paul Turnbull at the ANU’s Centre for Cross-Cultural Research (1998–2002) with the support of the National Library of Australia. In collaboration with Philippa Martyr of the University of Tasmania, and Alan Mayne of the University of Melbourne, Turnbull created among the earliest Australian online projects in the humanities, ‘Pictures of Health’ and the ‘History Info-server’ (1996).

[SLIDE: Paradisec]

Linda Barwick and Nick Thieberger established Paradisec (the Pacific and Regional Archive for Digital Sources in Endangered Cultures) in 2003. The archive began with an original focus on the Asia-Pacific region but has since expanded to include materials from all over the world. Its ongoing function is to digitise and archive records that relate to cultures and languages that may be under threat, thus providing citable research data as well as access for speakers of the languages. Its primary motivation is to ‘make field recordings available to those recorded and their descendants’. Supported by a consortium of three universities (University of Sydney, University of Melbourne, and the Australian National University), with over eleven hundred languages included to date, in 2016 the resource was given a Special Commendation by the UK’s Digital Preservation Awards.

[SLIDE: Australian Dictionary of Biography]

The *Australian Dictionary of Biography* (ADB; 1966–, General Editor Melanie Nolan) is the premier reference resource for the study of the lives of those who were significant in Australian history. A continuous publication project for more than fifty years that has produced eighteen volumes in print, ADB was launched as a freely accessible online resource in 2006. The database offers sophisticated searching and visualisation functions enabling users to explore networks of association between ADB subjects and other individuals, places and events referred to in companion sites including Obituaries Australia, Indigenous Australia, Women Australia, and Labour Australia under the shared framework of People Australia. The ADB initially utilised the Online Heritage

Resource Manager (OHRM) database system developed by the Australian Science and Technology Heritage Centre (Austehc) at the University of Melbourne, but since 2011 has been hosted independently at the Australian National University.

Numerous other projects have been initiated by Gavan McCarthy at the University of Melbourne via the Australian Science Archives Project (1985–1999), Austehc (1999–2006) and, since 2007, by the eScholarship Research Centre, employing the OHRM and other software platforms. These projects have included the 'Encyclopedia of Australian Science', 'Encyclopedia of Melbourne', 'Australian Women's Register' and the 'Find & Connect Web Resource' (jointly with the Australian Catholic University). The 'South Seas Online' project referred to above was also built on the OHRM.

[SLIDE: Design and Art Australia Online]

Another premier reference resource and biographical project with a long history, 'Design and Art Australia Online' was launched in 2007. It drew content from the *Dictionary of Australian Artists*, which was first published in 1984 but had its origins in the 1970s. It now includes a wide range of resources relating to 'Australian artists, designers, craftspeople and curators with information about artworks, event histories and collection details from major cultural institutions and individual researchers'. A large-scale collaboration, the online project was initially led by Vivien Johnson, and from 2010 by Ross Harley.

[SLIDE: Midja project]

e-Research and digital humanities have had a strong connection in the Australian context. This is demonstrated through the digital projects of the eResearch lab at the University of Queensland led by Jane Hunter since 2005, addressing topics including Indigenous knowledge management; digital preservation services; annotation tools for 3D artefacts; electronic scholarly editing tools; and knowledgebases for art conservators. More recently, this group has also been focusing on e-Social Science, analysing Australian Bureau of Statistics and urban transport data, developing dashboards for 'smart cities' and investigating relationships between housing, education and health in Indigenous communities, through the Midja project.

[SLIDE: Trove]

In Australia, as in other parts of the world, libraries have played a pivotal role in supporting the evolution of digital humanities as a field and set of activities that spans academia and the GLAM (Galleries, Libraries, Archives and Museums) sector. National and state digitisation programs led by collecting institutions are increasing researcher and public access to a wealth of otherwise hidden resources.

One of the best-known national digital projects is 'Trove'. With its genesis in the late 1990s, Trove was formally planned in 2008 as a portal to the National Library of Australia's discovery services and described in 2010 as 'a revolutionary new free search engine'. Trove has grown to become a full-text repository resource providing access to a vast array of information about Australia and Australians, an aggregator of diverse digital content and also a highly successful crowdsourcing platform for the correction of OCR-digitised content of Australian newspapers.

Making connections between digitised collections is one of its main purposes of Trove, which has become one of the most well-used publicly funded websites in Australia.

Trove is only one example of a growing national and international trend towards developing research infrastructure that aggregates material from multiple collections. I will discuss this further with reference to a selection of Australian examples.

[SLIDE: aaDH 2018 conference]

It was appropriate that the fourth conference of the Australasian Association for Digital Humanities, held in 2018, was on the theme of 'Making Connections'. This conference highlighted one of the most important and growing areas of digital humanities, which is the development of standards and systems to aggregate data from multiple sources.

Australia's research community relies largely on government funding, with a centralised approach that suits the country's very large land size and relatively small population. Government agencies have planned, on a national scale and over the course of more than a decade, a set of interlinking and mutually supportive infrastructures that address multidisciplinary requirements, ranging from shared data storage to metadata standards, discovery tools, supercomputing and cloud services. Achieving efficiency at scale and avoiding duplication of effort and resourcing are clear drivers in the Australian context, as they are internationally.

[SLIDE: Humanities Networked Infrastructure]

In Australia the development of national infrastructure designed to bring together and link major digital humanities projects and online collections began around 2012 when the first major project was funded by the Australian government. This project was the Humanities Networked Infrastructure (or HuNI as it is known), a national service for the study of Australian culture and history that I was involved with developing. Launched in 2014, HuNI was in development at the same time that plans were being made to establish an Australasian digital humanities association. These two significant initiatives provided mutual support and together helped to build momentum and visibility for the digital humanities field.

Data aggregation and integration have been a focus in many recent large-scale digital humanities infrastructure projects worldwide, including HuNI. A key challenge for many of Australia's major digital collections and projects (some of which have been described in this presentation) has been that these collections and resources are dispersed across multiple locations, institutions and agencies, and they mostly take the form of stand-alone, subject-specific repositories with very different information architectures.

[SLIDE: HuNI data model]

HuNI's data model is structured around six core entities of: concept, event, organisation, person, place, and work. This project is an example that demonstrates how digital linking and aggregation of disconnected repositories can produce a resource that is much bigger than the sum of its parts. HuNI aggregates data from 31 of Australia's most important humanities and creative arts datasets and makes them available for use by researchers across the arts and humanities and more widely by the general public (including through an Application Programming Interface [API], and as Linked Open Data, with a set of tools to work with the data). HuNI's main objectives are to make Australia's wealth of cultural resources more accessible and connected; to break down barriers between humanities disciplines and support cooperation and data sharing, nationally and

internationally; to create efficient workflows for working with cultural data centred around enhanced discovery, analysis and sharing; and for the HuNI data aggregate service to lay the foundation for collaborative cross-disciplinary online research capability into the future. By providing easy access to these rich repositories, HuNI is raising awareness and stimulating interest in Australian culture and history.

[SLIDE: Alveo virtual laboratory]

A further example of an Australian infrastructure project that has provided direct support to the digital humanities field is the government-funded 'Alveo virtual laboratory', for human communication science, which was funded around the same time as HuNI, and has proved to be a very important resource for linguistics and related research.

[SLIDE: TLCMap]

A final, more recent example, I would like to highlight, is the Time-Layered Cultural Map project, coordinated by Hugh Craig at the University of Newcastle, in which I am also involved as an investigator through Edith Cowan University. This project also has the goal of mass data aggregation, but in this case it is focused on providing access to these resources based on discovery and analysis via geo-spatial data of a very wide variety of kinds. The project is just starting, but it is an online research platform to deliver researcher driven national-scale infrastructure for the humanities, focused on mapping, time series, and data integration. The TLCMap will expand the use of Australian cultural and historical data for research through sharply defined and powerful discovery mechanisms, enabling researchers to visualise hidden geographic and historical patterns and trends, and to build online resources which present to a wider public the rich layers of cultural data in Australian locations. TLCMap is not a singular project or software application with a defined research outcome, but infrastructure linking geo-spatial maps of Australian cultural and historical information, adapted to time series and will be a significant contribution to humanities research in Australia. For researchers, it will transform access to data and to visualisation tools and open new perspectives on Australian culture and history. For the public, it will enable increased accessibility to historical and cultural data through visualisations made available online and in print.

[SLIDE: title slide]

The second part of my talk has highlighted the role of foundational projects for the development of the digital humanities field in Australia with a focus on the benefits of data aggregation at various levels and scales. As our everyday lives are increasingly influenced by aggregation of data, so too are research practices.

These trends in digital humanities point to a future in which researchers will be able to utilise comprehensive global data from many different sources and countries, and across languages, to form diverse and inclusive infrastructures and methods for data sharing and analysis. This will further advance knowledge and understanding and will develop new focus areas for digital humanities research as the field continues to expand and as digital technologies make an even greater impact in society and on research.