

Building Tools to Facilitate Data Reuse

Stefanie Kethers
Australian National Data Service

Andrew Treloar
Australian National Data Service

Mingfang Wu
Australian National Data Service

Abstract

The Australian National Data Service (ANDS) has been funded by the Australian Government since 2009, with a goal to increase the value of data to researchers, research institutions and the nation. To achieve this goal, ANDS has funded more than 200 projects under seven programs. This paper provides an overview of one of these programs, the Applications Program, which focused on funding software infrastructure to enable data reuse to demonstrate the value of making data available to researchers. The paper also presents some representative projects, a summary of what the program has achieved, and lessons learned.

Received 20 October 2015 ~ *Accepted* 24 February 2016 ~ *Revision Received* 29 June 2016

Correspondence should be addressed to Andrew Treloar, Monash University F610, PO Box 197, Caulfield East, VIC 3145

An earlier version of this paper was presented at the 11th International Digital Curation Conference Name.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution (UK) Licence, version 2.0. For details please see <http://creativecommons.org/licenses/by/2.0/uk/>



Introduction

ANDS was first funded by the Australian Commonwealth Government in January 2009, with the aim of transforming Australia's research data environment. Since that time ANDS has been working to enable the following four transformations:

- from data that are unmanaged to managed structured collections,
- from data that are disconnected to well-connected collections,
- from data that are invisible to researchers other than its creators to collections easily findable by other researchers, and
- from data that are single-use to reusable collections.

To enable the first three transformations, ANDS has been partnering with research and data producing agencies to make researchers and data holders more aware of the benefits of data sharing through setting up data management policies and procedures; to set up mechanisms for capturing data and metadata from e.g. instruments, computer simulations, sensors, historical data; and to set up research data asset registries and the discovery portal Research Data Australia (RDA)¹ - a national service for publishing data and making data citable and findable. These three transformations have supported the development of policies and procedures at research institutions for managing and sharing research data as well as putting in place data infrastructure to enable data sharing (Borgman, 2012; Tenopir et al., 2011).

However, all these are only means to an end – to make data available to researchers and beyond. Only when researchers are able to reuse data to advance human knowledge and to address some of the big problems faced by our society today will we see the true value of sharing data. To make this more apparent, ANDS set up the Applications Program, whose goal was to enable and promote the fourth transformation, producing compelling demonstrations of the value of having data available for re-use, and thereby addressing novel and complex research questions.

There are many factors affecting data reuse, such as the relevance of the data, the trustworthiness of data, the correct interpretation of data, and a range of technical issues (Howard et al., 2010; Faniel and Jacobsen, 2010). The Applications Program focused on the issue of bringing together data generated in different domains, and then providing data tools that enable or make it easier for researchers to discover and/or interrogate data, to construct productive and repeatable workflows, and to visualize data (Wu, Kethers and Treloar, 2013). This paper will present a range of projects funded under the Applications Program. These projects showcase the value of data reuse and the provision of data tools to researchers, but also to policy makers and the general public. These projects have been broadly divided into the following categories: Data Access, Workflows, Data Visualisation, and Connecting Science to Policy.

The paper is structured as follows: we will first lay out our motivation for setting up data tools to support research, which will be followed by an introduction of the Applications Program and exemplar projects from each the four categories described above. We will then present how the tools, services and output collections produced

¹ Research Data Australia: <http://researchdata.andis.org.au>

from the Applications projects are described in RDA. Finally, we will discuss challenges and lessons learnt from running the program.

The ANDS Applications Program

To demonstrate the value of making well-described data tools available to researchers, ANDS funded 24 projects through a program called Applications². Bioinformatics and Climate Change Adaptation-related research were the two main research areas covered by the projects (these clusters emerged over the process of project initiation); other projects dealing with urban planning, marine research, and public health were added to broaden the range of stories that could be told about the value of this approach. It was expected that these projects would result in data being transformed or integrated across multiple sources to produce new forms of information that enabled innovative, high-quality research outcomes. Projects were also encouraged to engage with other Australian National Research Infrastructure Capabilities such as NeCTAR (National eResearch Collaboration Tools and Resources)³, RDSI (Research Data Storage Infrastructure)⁴, TERN (Terrestrial Ecosystem Research Network)⁵, IMOS (Integrated Marine Observing System)⁶ and ALA (Atlas of Living Australia)⁷, so that a seamless data infrastructure would support the research, and the data tools would become more accessible and sustainable. Each project was scoped to have a one-year duration – most projects finished on time (excluding administration/contracting time), with a couple of projects taking about one and a half years to finish, mainly because of staff availability.

The core activity of an Applications project was to bring data from a variety of sources together, then to build tools to enable data reuse, or to make it easier for researchers to reuse the data. These data tools ranged from building new connections between data, or embedding new data analytic tools and algorithms, to visualising data to assist researchers in finding patterns among data. In some cases, a data tool would be an implementation of common computation or simulation models in a field so that researchers could all access these models, or an implementation of an existing workflow that would improve research efficiency.

The general development process of an Applications project started with high-profile champions (either research leaders or policy makers) in the selected areas looking at the available data, important research questions that they wanted to explore, and the needs of their research community. We targeted champions for two main reasons. First, these champions take an active role in shaping their discipline direction and were thus good candidates to promote a shared data infrastructure to their disciplines once they saw value in the respective projects. Second, these champions know very well what their respective community needs, but often lack resources to build such a data tools. The ANDS-funded Applications projects therefore intended to serve as a starting point after which further funding could be pursued, or a sub-community within the discipline could be formed to continue work on the data infrastructure.

² For a list of ANDS Applications projects, please visit: <https://projects.ands.org.au/getAllProjects.php?start=app> and <http://andsapps.blogspot.com.au/>

³ National eResearch Collaboration Tools and Resources: <http://nectar.org.au>

⁴ Research Data Services: <https://www.rds.edu.au/>

⁵ Terrestrial Ecosystem Research Network: <http://www.tern.org.au>

⁶ Integrated Marine Observing System: <http://www.imos.org.au>

⁷ Atlas of Living Australia: <http://www.ala.org.au>

After the champions set the project directions, the next step was to form a development team including project managers, software engineers and data analysts. Almost all of the projects adopted an agile software development methodology: a development team worked closely with champions or their nominated delegates to elicit, iterate and refine user needs throughout the application development process. In many cases, researchers involved in the project also invited their colleagues from other institutions or overseas to test an interim software version, and so got feedback from wider user groups within their research community.

We encouraged all projects to set up a blog or wiki⁸ to post about the software development progress, get feedback, and promote their application, and 20 out of the 24 projects did so. All projects were also encouraged to use an open sustainable source coding environment such as Google Code, GitHub or SourceForge etc., or to deposit all software development related resources (source codes, software installation guide and/or developer guide etc.) in an open source repository as a minimum requirement. Finally, projects were encouraged to promote what they had done through press releases, journal articles, conference papers and short videos. The videos, although mostly shot with minimal resources, have proven to be an effective way of letting researchers speak directly about the value of the projects.

The projects built a set of data tools to enable data transformation, data linkages and integration, data services, data analysis and modelling, data visualisation, and/or data manipulation workflows. It is difficult to classify these applications, as very often an application can cross many categories; even applications from the same categories may vary from discipline to discipline. Instead of giving a summary of applications from one category, we therefore present a few example applications.

A Data Access Example: Climate Model Downscaling Data for Impacts Research

A survey conducted by Tenopir et al. (2011) shows that nearly two thirds (67%) of researchers who participated agreed that a lack of access to data generated by other researchers or institutions is a major impediment to progress in science, especially in the areas of social science (80%) and environmental science and ecology (78%). Many factors may contribute to the difficulty of accessing data, such as data being under embargo, open only to a certain group of researchers, or not open at all. Furthermore, even open data may be too big to be accessed electronically as a whole. However, researchers may not need the whole data collection, but only a small portion of it for investigating their particular research question at a time.

Climate sciences use various climate models to study dynamics of the climate system and to project future climatic conditions. For example, the NARCLIM (NSW/ACT Regional Climate Modelling)⁹ collection is a petabyte-scale collection of Regional Climate Modelling simulation data, produced by climate change researchers from the University of New South Wales, Australia. The collection provides likely future climate change scenarios for New South Wales and the Australian Capital Territory. It includes a wide variety of climate variables at high temporal and spatial resolution, covers large regions on an irregular model grid, includes many climate variables and atmospheric layers, and is stored in NetCDF (Network Common Data Form) format.

⁸ See: <http://andsapps.blogspot.com.au/p/project-feed.html>

⁹ NARCLIM: <http://www.climatechange.environment.nsw.gov.au/Climate-projections-for-NSW/About-NARCLiM>

Traditionally, a collection like this has been used by climate researchers to compare and validate their models and to predict future climate change. However, such a collection can also be used by climate change impact and adaptation researchers to study and predict how climate change will affect our environment, society, and economy. Historically, it has been difficult for climate impact researchers to obtain downscaled climate data that can be more easily combined with data from impact fields. Some common problems include: an impact researcher may not understand the climate data output and implications of these data; a climate model output is usually very large, and an impact researcher may not have enough disk space to store the data, or to extract the data for a few relevant sites; an impact researcher's analytic tools may not be able to read the NetCDF data format, or to handle the irregular grid (Macadam et al., 2012). For example, when an agriculture business researcher wants to do farm scale planning, the researcher would like to access the NARCLIM collection, find the data related to that farm by locating the farm in the grid map, and then extract and derive relevant datasets.

One Applications project, Climate Model Downscaling Data for Impacts Research (CliMDDIR)¹⁰, built a data tool enabling a researcher to locate and extract a small portion of a dataset from a very large collection downscaled to their needs, i.e. the NARCLIM collection described above. Based on impact researchers' requirements, the CliMDDIR project implemented a user-friendly web portal that allows impact researchers to access impacts-relevant Regional Climate Model data. From the portal¹¹, impact researchers can extract subsets of data of selected variables and regions, re-grid or interpolate data to their selected regions, reformat the data into GIS, CSV and ASCII, calculate derived variables (e.g. pan evaporation), and apply statistical corrections, if necessary. A metadata record of an extract subset is also created along the way. Should the researcher think the subset is of potential interest to their colleagues, the researcher can choose to publish the metadata record to the data service portal or to Research Data Australia.

With the availability of climate modelling data, impact researchers and climate researchers can research our environment at a more complex and holistic level, and so enable research and collaboration across disciplines. Dr. Linda Beaumont from Macquarie University, an impact researcher working on the potential impacts of climate change on species and ecosystems, commented during the project that: "The impacts community has lacked consistent tools, with data being delivered in a variety of formats and at a coarse resolution that has often made them difficult to use. The CliMDDIR project will allow us to get our hands on up-to-date data and make more accurate assessments of climate change impacts. Currently, there is a substantial time-lag between when climate modellers develop data and when it becomes available to the impacts community in a useable format."¹²

A Scientific Workflow Example: Cancer Genomics Linkage Application

With research data becoming exponentially larger and more distributed, and data analysis increasing in complexity, more and more researchers are relying on scientific workflow management systems to conduct their data analysis. A scientific workflow management system can be an efficient tool to execute workflows and manage data sets

¹⁰ CliMDDIR: <https://web.archive.org/web/20160306154131/http://climddir.org/> (Note: the archive.org URL is being used because the climddir.org has been allowed to lapse).

¹¹ CliMDDIR Portal: <https://climatedata.environment.nsw.gov.au/>

¹² CliMDDIR project press release: <https://www.climate-science.org.au/content/158-climddir-project-website-portal-connected-software>

in various computing environments and to track provenance to make research more reproducible (Zhao, Raicu and Foster, 2008; Liu, et al., 2015). The second example Applications project demonstrated the automating of complex data processing and analysis tasks through re-usable workflows.

The Cancer Genomics Linkage (CGL) Application¹³ is built on the following use case: in responding to the International Cancer Genome Consortium (ICGC)'s initiative to catalogue the genetic changes of the 50 most common cancers, Professor Andrew Biankin and his colleagues from the Garvan Institute¹⁴ use genomics to understand the genetics of pancreatic cancer. Clinicians and biologists like Professor Biankin and his colleagues need to be able to analyse the raw genetic data, find genetic blueprints of each cancer patient, and make those genetic blueprints available to the cancer research community around the world, so that more researchers can access genetic blueprints, research and analyse them, and understand pancreatic cancer better. This will accelerate the process of finding treatments for this cancer.

However, the effective re-use of datasets has been limited by the ability of biologists and clinicians to access and use computational and data infrastructure. For example, the computational process for identifying and analysing genetic differences from the sequencing data generated by ICGC involves hundreds of steps and requires many computational tools, some of which have multiple versions. Access to genomic datasets of international importance and the ability to integrate them with the researcher's own clinical and genomic datasets are also critical in order to explore, discover and validate key genomic abnormalities that cause cancer. Traditionally, bio-informaticians with training in computer science, statistics, and engineering have written scripts to carry out this analysis process, but there is a critical need to put these tools into the hands of research biologists and clinicians.

The CGL Application provides integrated access to multiple data sources such as the ICGC variant database, or the DrugBank drug and drug target database, enables in-depth interrogation of cancer genomic data, and allows the comparison to other genomic data (Gorse, 2013). The application standardises some common analytical processes which are usually written by bio-informaticians, and turns the processes into Galaxy modules with the Galaxy server run by the Genomics Virtual Lab (GVL¹⁵). Galaxy, originally developed for computational biology, provides toolsets to build and assemble multi-step computational analyses into a workflow (Blankenberg, et al., 2010; Giardine, et al., 2005; Goecks, et al., 2010). The CGL Application also provides a mechanism to automatically record all aspects of an experiment. Some of this information is then extracted and published as a metadata record to selected data registries, such as Research Data Australia. This will allow for confidence in research repeatability and data re-use.

The integration of analysis tools, public and private datasets, and visualisation platforms in the CGL Application streamlines research and reduces the time from experiment to publication. Researchers such as Professor Biankin and his colleagues are able to access genomic datasets of international importance, and to integrate them with their own clinical and genomic datasets in order to explore, discover, and validate key genomic abnormalities that cause cancer, using user-friendly computational workflows¹⁶. Researchers can also publish and make available their analyses for re-use

13 Cancer Genomics Linkage Applications: <http://ap27-cgla.blogspot.com.au/>

14 Garvan Institute of Medical Research: <http://www.garvan.org.au>

15 Genomics Virtual Lab: <https://genome.edu.au/>

16 Hear Professor Andrew Biankin from the Garvan Institute discusses the importance of empowering clinician and biologist researchers through the use of bioinformatics workflows to help them fighting

by the community. Collaborations between researchers and across the community are enhanced through shared datasets, workflows, and customised toolsets.

A Data Visualisation Example: Brain Mapping National Resource

This third example presents a ‘big data’ visualisation tool. Modern scientific imaging systems such as x-ray radiography and magnetic resource imaging (MRI) generate massive volumes of data of up to 1TB per acquisition. Usually an image can only be viewed via specialized viewing software available at a scanning instrument, so that, even if data are made available, researchers cannot utilize the data on their own computers. Professor Charles Watson from Curtin University and Neuroscience Research Australia stated that when he and his collaborators started to work on high resolution MRI brain scans around 2007, they printed images of various sections of a brain and manually annotated the print-outs. These annotations were then sent back and incorporated into the original data set¹⁷. This presents a large barrier to collaborative research, as it is difficult to share data between sites and keep collaborators up to date with the latest observations.

The Brain Mapping National Resource project has implemented the web-based 3D dataset viewer TissueStack (Janke et al., 2013). TissueStack employs HTML5 technology that generates image tiles on the fly. With TissueStack, Professor Watson and his collaborators can view 3D micro-CT, MRI and re-stacked optical imaging (Histology) data, and collaboratively annotate remotely without the need to make a local copy of the data. TissueStack also enables researchers to slice images, present an image in any direction, and jump from one plane to another.

Professor Charles Watson, who collaborates with the Center for Advanced Imaging (CAI) at the University of Queensland and Duke University on different projects, commented that the ability to share data from the cloud, access it through TissueStack^{18,19}, is making a huge difference to the way we are able to interact, the ability for all participants to access the same dataset, to annotate it and to have a discussion on the way forward. The software TissueStack has been used by the Montreal Neurological Institute to display the BigBrain dataset²⁰, and by the University of Toronto to display large block-face mouse imaging datasets.

A Connecting Science to Policy/Public Example: POSITIVE PLACES - Spatial Analysis of Public Open Space

This final example integrates data from multiple sources into a coherent context, thereby augmenting the power of data in enhancing public engagement.

This project was set up to support researchers trying to understand how public open spaces (POS)²¹, including parks, reserves and bushland, affect public health, especially individuals’ and the community’s mental and psychological health. This understanding will help city planners create healthy, active, and sustainable communities and cities. To understand this problem space, researchers from the Centre for the Built Environment and Health (CBEH) at the University of Western Australia collected

cancer: <https://www.youtube.com/watch?v=jGY5GpWmJNg>

17 TissueStack - ANDS AP016 DOV video: https://www.youtube.com/watch?v=y-8u_0vY6MU

18 TissueStack is awesome: <http://tissuestack.blogspot.com.au/>

19 TissueStack: <http://www.tissuestack.org>

20 BigBrain LORIS Database: https://bigbrain.loris.ca/main.php?test_name=bigbrain&slice=3702

21 POSitive Places: <http://positiveplaces.blogspot.com.au/>

numerous datasets from local governments. Whilst many Local Government Areas (LGA) hold data on their parks, reserves, and other public open spaces, the quality and form of these data varies. Additionally, a lack of consistency in definitions, terminology and descriptions of POS has prevented comparative analyses.

Having identified needs from three user groups, i.e. the public, urban planners, and researchers, the project designed and implemented an easy-to-use web-based POS tool²², which provides members of the public with the ability to search and find details of parks in their local area. Specifically, the POS Tool allows a search by address, and the results provide details of the nearest five parks in their local area, as well as information on the amenities and facilities of those parks. The POS Tool offers the community a simple approach to finding consistent, up-to-date park information across the Perth metropolitan region, allows local governments to compare residential access to various types of park facilities within and between Perth's Local Government Areas and Suburbs, and enables better planning of land allocation and site location of parks as well as the allocation of park amenities relative to existing local services. Furthermore, planners can use the POS Tool to model future POS needs according to forecast and hypothetical demographic and population changes; researchers can use advanced features to upload the data of a defined region, and to test scenarios dealing with the relationships between changes in population structure for the defined area and the provision of parks (Bull and Boruff, 2013).

The POS tool shows the usefulness of the combined data sets not only to researchers but also to policy makers and the public. In Parks and Leisure Australia's Western Australia (PLA-WA) 2014 Awards of Excellence, the POS tool won both the Research and Use of Technology Categories. PLA-WA commented that: "[The] POS Tool provides planning professionals, across both private and government sectors in Western Australia with access to an extensive, accurate and unrivalled spatial database with a level of detail which has previously been unavailable for POS information in WA... and is fundamental to the effective rational planning for POS and its importance cannot be underestimated."²³

Challenges and Lessons Learnt

Through running the program, we have observed and learnt what worked and what did not work for our projects through the following means: 1) we ran two workshops early in the program to lay out expectations from both sides and to foster collaboration across projects, 2) each project gave a demonstration and submitted a final report at the end of the project which included lessons learnt and suggestion for future improvement, 3) we presented a poster (Kethers, Treloar and Wu, 2014) about the program at the eResearch Australasia 2014 conference with post-it notes on it to gather feedback from the wider community on sustaining eResearch effort as in these projects, 4) we interviewed six projects after the program finished about how these projects have been sustaining the efforts put into the projects. We group the lessons learnt into the following categories:

- Support from champions
- Funding/timeframes

²² POSitive Places portal: <http://www.postool.com.au>

²³ POSitive Places blog: <http://positiveplaces.blogspot.com.au/search?updated-min=2014-01-01T00:00:00-08:00&updated-max=2015-01-01T00:00:00-08:00&max-results=1>

- Staff skills and knowledge
- Outreach / community building
- Sustainable data availability

Support from Champions

As introduced above, each project was required to nominate one or more champions that were either influential researchers or policy developers, with a strong interest in the project outcomes. Most projects had at least one champion who played an active and important role in promoting the data tool to their relevant community and ensured that the tool met their community's needs. Often, these champions also actively sought further funding to enhance the tool. Projects driven by active champions who know and can articulate the requirements of their community for the tool seemed to improve uptake. However, it is important that the champion(s) stay involved throughout the project. In particular, it seemed that tools tend to stay in use longer if the champions driving the project help to grow a community of data providers and consumers around the tool.

Funding/Timeframes

The projects were funded for 6-15 months. Due to this limited time frame, the projects generally focused on developing what was required for the purpose – there was no time to do a more generic modular design. For example, while it would have been useful to combine effort across several of the Applications projects early on (e.g., all 'map' projects pooling resources to build one tool), this did not happen due to different preferences, timeframes, and priorities. Because of the short duration of the available funding, there was also often no staff continuity, which in turn meant that software will invariably be wholly or partially rewritten once more resources become available. Furthermore, tools and services need to evolve, which requires ongoing commitment, but that was not reflected in the funding situation.

Staff Skills and Knowledge

Staff expertise and preferences are crucial – and can sometimes limit sustainability. For example, a developer may choose a less popular programming language because of their expertise in the language, and this code may have to be re-written later to ensure compatibility with mainstream applications. Developers and providers also need to know who does what in the community, and where to get help. This information is not always obvious.

Most of the Applications project teams had software developers as key team members, some also had data managers. Over the course of their project, project staff (including project managers) gained expertise in a wide range of eResearch issues, especially on research data management such as DOI minting (Klump et al., 2015) and data licensing. Some of them are now bringing this skill and knowledge set into their new data-oriented projects.

Outreach/Community Building

As discussed above, 20 out of the 24 projects set up a blog or wiki to communicate their software development goals and development progress, get feedback, and promote their data tool. Those who kept their blogs up to date had a positive attitude towards such an approach; these regular updates also served as development journey log for later reference. However, for such outreach activities, developers and project managers would have appreciated help from more ‘commercially-minded’ stakeholders. One suggestion we received was that the funder could help write blog posts and other promotional material based on conversations with the project team, rather than let technical people struggle with describing the tool from a different perspective.

Presentations and talks at various conferences and workshops helped to increase the publicity of the projects and to generate potential interest in utilizing the data tool. This extended the reach of the project beyond the interest raised through existing contacts in the community.

Sustainable Data Availability

A fundamental component of the Applications projects was data. In many of the projects, data came from government agencies at various levels (Federal, State or Council). These agencies needed to have a plan to make data open and keep it up-to-date. Sometimes, the agency may need to invest extra effort to digitalize and/or de-sensitise data, yet this is not always a priority, and agencies often lack the resources or the capability to do so.

Discussion

The ANDS Applications program brought IT professionals and data analysts together with researchers to build data tools to enable and support new ways of research. We have seen that many of the researchers engaged in the projects truly appreciate the new (or improved) research they can do with shared data through the data tools developed. Most of the projects tell compelling stories (as video clips, blog posts or media releases, and by winning awards and further grants etc.) of the benefits of building tools to make data reusable. Through the journey with our partners, we have seen many challenges ahead of us in order to reduce barriers between researchers and data. For example, many applications have similar components for data transformation, data integration and data visualisation. A further step would be to make these tools more generic so that they can be easily adopted by more research communities other than the initially targeted ones. Furthermore, although the research community was the primary target of the Applications projects, we have seen that many data tools can also support policy makers and the general public to make informed decisions. In the future, when we build data tools, we may need to think out of the research box, to better connect research with policy makers and the public through shared data and data tools, so the government and the general public can see an even greater value in their investment in research. The final challenge is to work out how to meet the needs of a range of user communities wider than those engaged in this project in a way that is fundable and leads to sustainable outcomes.

Acknowledgements

We would like to thank former ANDS staff member Dr. Jeff Christiansen for his efforts in shaping and managing some of the Applications projects, and all our partners, project members and research champions for their contributions to the Applications program.

References

- Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., Taylor, J. (2010). *Galaxy: A web-based genome analysis tool for experimentalists*. *Current Protocols in Molecular Biology*. 2010 Jan; Chapter 19:Unit 19.10.1-21. doi:10.1002/0471142727.mb1910s89
- Borgman, C.C. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6):1059-1078.
- Bull, F. & Boruff, B. (2013). *POSitive places: New data, tools and approaches for examining the role of public open space in health and planning research*. Presented at the Association of American Geographers, Los Angeles, California.
- Faniel, I.M., Jacobsen, T.E. (2010). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work (CSCW) 19(3-4)*, 355-375. doi:10.1007/s10606-010-9117-8
- Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W.J., Nekrutenko, A. (2005). Galaxy: A platform for interactive large-scale genome analysis. *Genome Research* 15(10):1451-5. doi:10.1101/gr.4086505
- Goecks, J., Nekrutenko, A., Taylor, J., & The Galaxy Team. (2010). Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* 11(8):R86. doi:10.1186/gb-2010-11-8-r86
- Gorse, D. (2013). *A genomics virtual lab for cancer research*. Bio-IT World Asia, Singapore. Retrieved from <http://slideplayer.com/slide/4640604/>
- Howard, T., Darlington, M., Ball, A., Culley, S., & McMahon, C. (2010). *Opportunities for and barriers to engineering research data re-use*. ERIM Project Document erim3res100805tjh10. Bath, UK: University of Bath. Retrieved from <http://opus.bath.ac.uk/21166/1/erim3res100805tjh10.pdf>
- Janke, A., Lin, M.K., Nicolini, O., Waxenegger, H., Ullmann, J., & Galloway, G.J. (2013). TissueStack: An open source HTML5 web based imaging viewer. *Journal of Frontiers in Neuroinformatics*. doi:10.3389/conf.fninf.2013.09.00120

- Kethers, S., Treloar, A., & Wu, M. (2014). *Sustaining eResearch efforts: Some lessons learnt*. Poster presented at e-Research Australasia 2014. Retrieved from <https://drive.google.com/file/d/0B3urY1fwzAEcRXRyMnlWcldPMIU/view?usp=sharing>
- Klump, J., Huber, R., & Diepenbroek, M. (2015). DOI for geoscience data - how early practices shape present perceptions. *Journal of Earth Science Informatics*, pp.1-14. doi:10.1007/s12145-015-0231-5
- Liu, J., Pacitti, E., Valduriez, P., & Mattoso, M. (2015). A survey of data-intensive scientific workflow management. *Journal of Grid Computing*, pp.1-37.
- Macadam, I., Fuchs, D., Hotz, A., & Pitman, A. (2012). *Climate model downscaling data for impacts research: An ANDS initiative*. Poster presented in the e-Research Australasia Conference. Sydney, Australia. Retrieved from https://eresearchau.files.wordpress.com/2012/10/eresau2012_submission_155.pdf
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 6(6). doi:10.1371/journal.pone.0021101
- Wu, M., Kethers, S., Treloar, A. (2013). *Getting from managed to reused: Making it easier for researchers to do something useful with data*. Proceedings of e-Research Australasia 2013, Retrieved from http://eresearchau.files.wordpress.com/2013/08/eresau2013_submission_49-2.pdf
- Zhao, Y., Raicu, I. & Foster, I. (2008). *Scientific workflow systems for 21st century, new bottle or new wine?* In Proceedings of the 2008 IEEE Congress on Services – Part I. pp. 467-471.