

Data and Services Discovery projects - Transformative Data Collections

Title

TD96 Enhancing the data, tools and policy translation resources on 'CARDAT'; the NHMRC Centre for Air pollution, energy and health Research (CAR , www.car-cre.org.au) Data and Analysis Technology (DAT) platform.

Approach

How did you deliver on the proposed approach as outlined in your original application?

- What activities did you undertake?
 - Data publishing, new tools developed and a conceptual modelling workshop.
- Which participants or collaborators were involved?
 - In the workshop there were three policy workers from environmental health branch in the EPA Victoria, one environmental epidemiologist and policy worker from University of Sydney, one data scientist from UQ and two Geospatial Health data portal workers from University of Canberra. In the data publishing component, we met with the Product Manager Cloud Services at Cloudstor (AARNET) and developed the system design for long-term data storage.
- Were any outputs produced, e.g. data collections, publications, reports, presentations, software? Include copies, links or DOIs.
 - Yes, datasets and tools were made public (see below).
 - An executive overview of the IT infrastructure was prepared, see attached (CARDAT_Infrastructure_Report_20191008.pdf).
 - A workshop report was prepared, see attached (Translational-science-project-workshop01-summary_2019-10-08.pdf).
 - A new website hosting our data inventory (<https://cardat.github.io>)
 - The FAIR spreadsheet assessment of our data platform (see attached).

This project enabled us to make some of the non-restricted environmental health datasets and tools within CARDAT a public resource for all Australians. The keyword is "non-restricted" as this means it is only data and tools that we can legally share based on our agreement with the data/tool owners. During October a full searchable listing of all our datasets was made available on our new website <https://cardat.github.io>, and requests for access are currently possible by contacting our new generic email address car.data@sydney.edu.au.

We have begun making available many of the outputs of our models, as well as the value-added

general data (e.g. improvements made to ABS or AIHW datasets) unrestricted. We have made a new data publishing procedure utilizing both the Cloudstor Service from AARNET and the Open Science Framework (OSF). Two examples are:

- Hanigan, I.C. (2019) Drought by 2016 Local Government Areas. Derived from ABS 2016 LGA data, and AWAP grid rainfall data. Centre for Air pollution, energy and health Research. <https://cloudstor.aarnet.edu.au/plus/f/3711709620> and <https://osf.io/5wq3f/>
- Hanigan, I.C. and Geromboux, C. (2019). ABS SA2 2016 combined with AWAP weather data for 1991-2018. . Centre for Air pollution, energy and health Research. <https://cloudstor.aarnet.edu.au/plus/f/3885564812> and <https://osf.io/sf9kg/>

Which adds to datasets made public during the previous two years of CARDAT developments:

- Hanigan, I.C. (2018), DROUGHT_BOM_GRIDS_1890_2008, Centre for Air pollution, energy and health Research. The Australian Centre for Air pollution, energy and health Research (CAR). <https://cloudstor.aarnet.edu.au/plus/f/1864605413> and <https://osf.io/np5xd/>
- Hanigan IC, Morgan GG, Williamson GJ, Salimi F, Henderson, Sarah B., Turner MR, Bowman DMJS, Johnston FH (2017) The Biomass Smoke Validated Events Database. Centre for Air pollution, energy and health Research. <http://dx.doi.org/10.4225/64/59276c30b3a01> and https://github.com/swish-climate-impact-assessment/biomass_smoke_events_db

We are working toward making more of our datasets and tools available under this new procedure.

There will be some restricted data outputs which will be mediated by our data management procedures overseen by the CAR Data Scientist. Restricted data will be those that are potentially identifiable, due to conditions imposed by data custodians or involve intellectual property rights and made available under mediated data sharing agreements and licensing.

FAIR

How has your project enabled making data more [FAIR](#)?

- Please attach the FAIR assessment spreadsheet to this report

Collaboration and coverage

What level of collaboration and national coverage did you achieve?

- Has the data collection become more national as a result of this project in terms of:
 - Contributors = new link with systems thinking experts Barry Newell and Katrina Proust (ANU)
 - Collaborators = new links with EPA Victoria and strengthened links with UC
 - Governance = utilizing the OSF has increased the national (and international) accessibility of the project

Sustainability

How will the project outcomes be sustained?

- What agreements are in place to sustain the outcomes of the project?
 - The Centre for Air pollution, energy and health Research (CAR) is an NHMRC funded Centre of Research Excellence (CRE) funded from 2017 to 2021. In 2019 CAR has funded a team of one data scientist and one data curator to build the CAR Data and Analysis Technology (CARDAT) data platform which comprises a data portal (based on Cloudstor) and remote desktop workspaces (based on the NCRIS EcoCloud service CoESRA).
 - Our meeting with AARNET Product Manager Cloud Services Gavin Kennedy covered the details about the CloudStor service offering, as well as the RoadMap of where the service is heading over the next few years. We are confident that the agreements governing our access to the storage solution offered by AARNET is sustainable in the medium to long-term future.
- What are the existing ingredients that enable sustainability?
 - NHMRC funding till 2021
 - AARNET Cloudstor support via merit allocation
 - NCRIS NeCTAR node
 - The data platform is used by multiple institutions and agencies
- What steps will you take beyond the project to sustain the data collection and/or outcomes of the project?
 - The CAR support is committed for a data scientist position in 2020. It is hoped that in 2021 we will leverage off the success of the data platform to enable future funding of support staff.
 - We have collaborated in two new funding applications related to this work
 - 1) The latest call for ARDC funding to support platforms (Title: “Multi-disciplinary Scalable Research Virtual Desktop Platform”, Lead Investigator S. Guru, UQ).
 - 2) UKRI-NHMRC international collaboration grants (Title: Healthy, Equitable and Sustainable Urban Mobility to Prevent Non-Communicable Diseases in the UK and Australia, Lead Investigator S. Vardoulakis, ANU).

Learnings

Were there key lessons learned in bringing together a Transformative Data Collection?

- Collaborative working with people geographically separated via Zoom and the Remote Desktop data analysis platform was a great approach and worked well
- Re-use and adaptation of existing approaches (such as the collaborative workshop we used) can be fruitful

What does the ARDC need to address at the national level to make the process easier for growing other Transformative Collections?

- Multiple communities with different perspectives (i.e. policy workers versus data analysts) can come up with sometimes under appreciated insights about each other's worlds

Impact

What are the research outcomes impacted by this project?

- What research publications or grants has/will this collection enable?
 - The project was informed by the Usyd and Vic EPA had a paper published and new paper ideas have been generated that build on the new possibilities offered by this project.
- What new collaborations and/or communities has this project enabled?
 - Now that components of our data collection are public, we have many opportunities.
 - Our collaboration style with EPA Vic policy makers and UC Australian Geospatial Health Hub has been renewed and reinvigorated.
 - New mediated access procedures have enabled a new NSW government planning department to access our data
- What new research projects or programs have been enabled?
 - Standardised methodologies for environmental health impact assessments (HIAs) with EPA Victoria and the NSW Ministry of Health
- Will the project enable new research areas or approaches?
 - Later in 2019 we will use this project to support a new workshop on a case study in air pollution health impact assessments and regulation as an example. The example of translational science will be drawn from the real-life example of the co-driven development of health impact assessments for alternative air pollution exposure control strategies by environmental epidemiologists, public health and urban planning stakeholders.

What is the broader anticipated impact of this project, beyond the organisation and/or research sector?

- Who or what might benefit from the results of the project (industry, community, government, wider public, etc)?
 - Policy workers and their relationship with government and through that the public
 - Better and more evidence-based policy discussions
- What is the anticipated nature of the impact, including social, economic, cultural, and environmental impacts?
 - The impact is potentially very broad. Certainly, the air pollution control example has major implications for health because the threshold of allowable pollution may come done
- What is the expected extent of the impact and the time period in which it may occur?
 - Nation Clean Air Agreement 2020 and beyond (e.g. next phase of NEPM)

- Will you put in place pathways to ensure future impact?
 - We will engage our policy worker community so that we know their deadlines for the relevant policy discussions
 - We will complete the next workshop around an applicable policy framework in a timely fashion

Report prepared by: Ivan Hanigan PhD,
University of Sydney School of Public Health
Date: 08/10/2019