



Moving data around: integrated research workflows for curating, and publishing data

Maude Frances (1), Steven McEachern (2), Daniel Bangert (1), Carolien van Ham (1), Janet McDougall (2), Luc Betbeder-Matibet (1) UNSW Sydney
 (2) Australian Data Archive, Australian National University

IASSIST 2017, Lawrence, Kansas, 23-26 May

Fast facts

<p>Founded 1949 Located in Sydney</p>	<p>8 Faculties Art & Design Arts & Social Sciences Built Environment Business Engineering Law Medicine Science</p>	<p>\$215m invested in campus infrastructure across 123 projects</p>
<p>54,043 STUDENT ENROLMENTS</p>	<p>4,200 wireless stations on campus and tripling of network speed</p>	
<p>14,392 INTERNATIONAL STUDENTS</p>	<p>19,777 COMMENCING ENROLMENTS</p>	<p>1 UNIVERSITY COLLEGE - UNSW Canberra at the Australian Defence Force Academy</p>
<p>6,228 STAFF</p>	<p>279,226 ALUMNI</p>	<p>47 125 18 Schools UNSW & external centres and institutes On-campus colleges and affiliated student accommodation</p>



Overview

- Researcher perspective: current practice and needs
- Research infrastructure project for curation and publication of data
- Infrastructure and workflow
- Benefits
- Next steps



Current practice: curation and publication

- Most of my colleagues publish their datasets on their own (non-institutional) websites, along with documentation such as the codebook and publications that used the data
- Larger group efforts for data-collection publish via own institutional website, i.e. Varieties of Democracy, Polity IV, Freedom House
- Some also deposit datasets at data archives like Harvard Dataverse.
- Need to balance ***recognition, impact, and efficiency.***



Data on Elections, Democracy and Autocracy (DEDA)

- The Quality of Elections in Democracies and Autocracies dataset. Election and election fraud in over 900 elections in 97 third and fourth wave regimes between 1974-2012, including data on presidential and parliamentary election timing, results, electoral fraud, characteristics of opposition and incumbent, turnover, and presence of civil wars and coup d'etat.
- Developed for PhD (van Ham 2012)
- Currently being expanded until 2016 and to over 120 countries, funded by the Australian Research Council DECRA "Getting Elections Right", RG142911.

4



Research

I am a Lecturer in Comparative Politics at the School of Social Sciences at UNSW and an ARC Discovery Early Career Research Award recipient. My research interests are: democratization and authoritarianism, election integrity and electoral fraud, and political representation, accountability and legitimacy.

I am currently working on the following research projects:

- **ARC DECRA project 2015-2017 Getting elections right? Electoral reform in democracies & autocracies.** Project investigates: (a) why and under what conditions electoral reform in electoral democracies and electoral autocracies is successful; (b) how election integrity can be effectively strengthened; and (c) how election integrity affects democratization. This project involves mixed-method research using large-N quantitative analyses as well as qualitative country-case studies. The research project is connected to the Electoral Integrity Project at Sydney University and the Varieties of Democracy Project at Gothenburg University.
- **ANDS project 2016-2017 Data on Elections, Democracy and Autocracy (DEDA).** Website hosted at UNSW mapping quantitative datasets on Elections, Democracy and Autocracy. Website: (a) enables sharing of data on these topics; (b) provides a macro data guide to other important datasets in this field accompanied by syntax to merge datasets to enable easy use; (c) aims to provide online data visualisation tools that would allow researchers and the general public to carry out simple data analyses online. Data on Peace & Conflict and Human Rights to be added in 2nd stage of project.
- **Electoral Management Research Network 2016-ongoing. Improving Electoral Management: The Organisational Determinants of Electoral Integrity,** with T. James, H. Garnett and L. Loeber, University of East Anglia and McGill University. First phase of the project is to collect data on EMB organizational structure and personnel through two surveys set out among EMBs in 30 European countries with support from the Venice Commission. Surveys currently in the field, phase 1 to be completed in April 2017, and results collated in a series of journal articles as a special issue. Phase 2 consists of setting up the Electoral Management Research Network, bringing together European and international scholars in this field to strengthen collaboration and dissemination of research. Funding from University of East Anglia, EIP project and UNSW.

5



ANDS High Value Collection Program



UNSW partners: School of Social Sciences, Office of PVC – Research Infrastructure, UNSW Library

External partner: Australian Data Archive, Australian National University

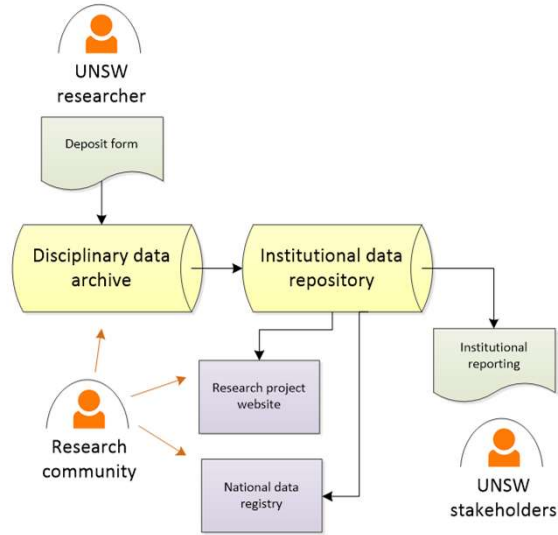


Goals of the project

1. Curate and openly publish several datasets on elections, democracy and autocracy collected by the researchers and colleagues at School of Social Sciences (SoSS) UNSW.
2. Create an institutional website on data on Elections, Democracy and Autocracy
 - enables sharing of data collected by researcher and UNSW colleagues on these topics
 - includes a macro data guide to other important datasets in this field
 - guides users to online data visualisation tools that would allow researchers and the general public to carry out simple data analyses online.



Data curation and publishing workflow



AUSTRALIAN DATA ARCHIVE
DATA DEPOSIT INFORMATION FORM

PAGE

Section One: Title and Abstract

Study Title

Quality of Elections in Third and Fourth Wave Regimes 1974-2009


Study Abstract

The dataset Quality of Elections in Third and Fourth Wave Regimes 1974-2009 collects data on elections and election fraud on 885 elections in 97 third and fourth wave regimes between 1974 and 2009. The dataset is in country-year format and includes data at the country-level and at the election level.

At the country-level, the dataset includes indicators on region, year of independence, previous democratic experience, civil war, coup d'etats, regime breakdown as well as indicators on whether the regime held elections with de jure multi-party competition and de jure universal suffrage.

At the election level, the dataset includes indicators on presidential and parliamentary elections, constituent assembly elections, election results, vote and seat shares of incumbents and

Dataset in ADA Dataverse

 Dataverse

[Q](#)
[About](#)
[Guides](#)
[Support](#)
[Steven McEachern](#)

DEV ADA Dataverse > **Quality of Elections in Third and Fourth Wave Regimes 1974-2009**

Metrics
0 Downloads
Share
Download
Publish
Edit

Quality of Elections in Third and Fourth Wave Regimes 1974-2009 Draft Unpublished

van Ham, Carolien, 2017, "Quality of Elections in Third and Fourth Wave Regimes 1974-2009", doi:10.5072/82/N1CGES, DEV ADA Dataverse, DRAFT VERSION Cite Dataset


[Learn about Data Citation Standards.](#)

Description The dataset Quality of Elections in Third and Fourth Wave Regimes 1974-2009 collects data on elections and election fraud on 885 elections in 97 third and fourth wave regimes between 1974 and 2009. The dataset is in country-year format and includes data at the country-level and at the election level. At the country-level, the dataset includes indicators on region, year of independence, previous democratic experience, civil war, coup d'etats, regime breakdown as well as indicators on whether the regime held elections with de jure multi-party competition and de jure universal suffrage. At the election level, the dataset includes indicators on presidential and parliamentary elections, constituent assembly elections, election results, vote and seat shares of incumbents and opposition, turnover, and the quality and integrity of elections. The dataset and research based on these data is described in-depth in van Ham (2012) and van Ham (2014).

Subject Social Sciences


Keyword Elections

Related Publication Van Ham, C. 2014. Getting elections right? Measuring electoral integrity. Democratization. doi: 10.1080/13510347.2013.877447



Metadata

ADA form	DDI	ResData	RIF-CS	DataCite
Study title	titl	dc:title	name[TYPE=primary] > namePart	title
Study abstract	abstract	dcterms:abstract	description[TYPE=full]	description[TYPE=abstract]
Principal investigator	AuthEnty	foaf:firstName foaf:lastName	citationInfo > citationMetadata > contributor > namePart	creatorName
Geographic coverage	geogCover	core:GeographicLocation > ms21:spatialInformation	coverage > spatial[TYPE=text]	geoLocation > geoLocationPlace
Time period covered	timePrd	dcterms:PeriodOfTime > core:startDate core:endDate	coverage > temporal > date[TYPE=dateFrom] [TYPE=dateTo]	
Time dimensions	timeMeth	... <i>Not collected</i> ...	dates[TYPE=created] > date[TYPE=dateFrom] [TYPE=dateTo]	date[DATE=Collected]
Funding agency	fundAg	ms21:organisationName		fundingReference > funderName
Distributor	distrbr	... <i>Not collected</i> ...	citationInfo > citationMetadata > publisher	Publisher
Related publications	relPubl	ms21:hasRelatedPublication	relatedInfo[TYPE=publication]	relatedIdentifier[relationType=IsReferencedBy, IsDocumented By]



ResData

About ResData

ResData is a Library service for managing UNSW research data. The service has two main components:

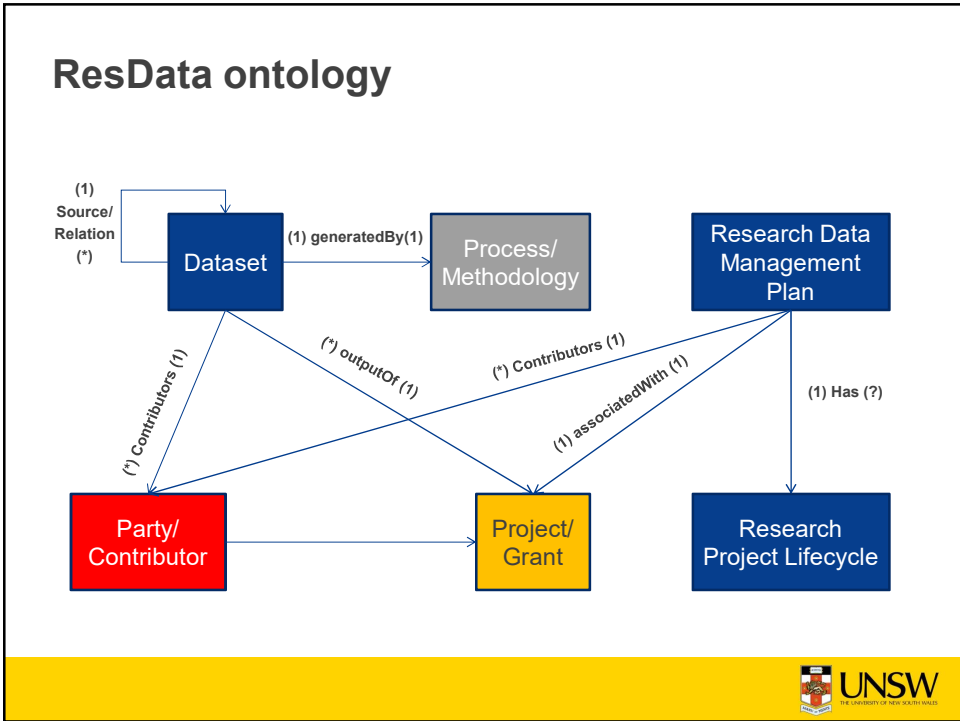

- Research Data Management Plan (RDMP) - allows UNSW researchers and HDR Candidates to record metadata about research projects and associated research data. Completion of an RDMP is a requirement for allocation of storage in the UNSW Data Archive.
- A catalogue of UNSW datasets and collections of research materials.

Catalogue records in ResData are published to Research Data Australia, a discovery service for Australian research data. The service is supported by the Australian National Data Service (ANDS) through the National Collaborative Research Infrastructure Strategy Program and the Education Investment Fund (EIF) Super Science Initiative, as well as through UNSW Australia.

Log in

Staff/Student ID:

Password (zPass):



Benefits

- Dataset recognized as research output through DOI
- Direct link to dataset at ADA allows publicizing dataset on my website, while data safely deposited, and findable via multiple pathways (ADA, website, library research collection)
- Increased usage of our data, not only by academics but also by students and practitioners/policymakers
- Increased visibility of UNSW research in these areas, supporting future grant applications and research collaboration with other universities in this field
- Provides a platform to expand the website in the future to include data visualization tools and data on peace & conflict, development, and human rights, providing an opportunity to showcase the key research strengths of the School of Social Sciences at UNSW
- Process of data depositing at ADA encouraged through documentation of data collection process and improved quality of codebook.

14



Recognition – Impact - Efficiency

- Recognition: Academic recognition for data collection effort
 - Usually via publications about the dataset or publications that analyse data
 - Data archive mints a DOI - dataset can be directly recognised as research output
- Impact: Platform for dissemination of data
 - Link to dataset from researcher's website - links directly to data archive source
 - Institutional repository links data to related publications and disseminates to national metadata registry
- Efficiency:
 - ADA submission process highly efficient: description of data & consent form.

15



What do we need to do next?

1. Shift from PDF based to direct entry of metadata into Dataverse (out of the box DV functionality)
2. Review of data and metadata by ADA staff
3. Publishing the data at ADA
 - Automated minting of DOI through Datacite (indirectly through ANDS minting service)
4. Harvesting metadata from ADA to UNSW ResData
 - a) OAI-PMH from Dataverse
 - Mapping already done from DDI to ResData
 - Need to configure harvester (in Dataverse)
 - OR
 - b) Harvest from Research Data Australia / Datacite