

## Impact Objective

- Strengthen data expertise and research partnerships between South Africa and the UK through the formation of a new network aimed at investigating access to, and use of, household energy data for policy-making and planning

# Building data infrastructure

*Louise Corti from the University of Essex talks about her recently completed project that formed a knowledge exchange and planning network which focused on addressing key policy challenges by sharing the scoping, planning and analysis of UK and South African household energy data*



**Do you consider current household energy data sharing practices to be effective and support current energy policy and planning?**

No. We need to use better data sources and techniques. Three research strands highlight this: firstly, studying UK domestic energy consumption to better inform strategies to deal with fuel demand and fuel poverty, and energy saving and efficiency; secondly, using longitudinal data to measure the impact of the roll-out of electrification in rural South Africa; and finally, utilising consumer data to study equity in household energy consumption in urban South Africa.

**What is the ultimate end value of the 'Smarter Household Energy Data' project?**

The most important outcomes are benefits for researchers and policy makers through access to high quality data, development of infrastructure for data intensive research and the knowledge exchange impacts that these bring. There is also much added value through capacity building in social science oriented 'data science', as well as the Newton Fund Impacts, which promote economic development and social welfare in South Africa.

**The project strongly relied on the collection of comprehensive big data sets, as well as the**

**techniques and software capable of analysing the data. Can you talk a bit about the different technologies developed for collecting and analysing the data?**

The UK Data Service (UKDS) wishes to modernise its data and analytics services to deliver and provide insight and intelligence to researchers in an efficient and cost effective manner. A solid technology framework is necessary to support large and complex data sets, meet security considerations for our spectrum of data access from Open, through Safeguarded to Controlled data, and deliver robust analytics. The UKDS chose open source technologies to avoid vendor lock in, including Hadoop and Apache Nifi, and worked with Cloudwick, Amazon Web Services (AWS) and Hortonworks, to successfully implement a hybrid solution to provide these features. The cloud allows us to perform on-demand scaling for search infrastructure and web traffic, and an on-premises infrastructure maintains tight governance controls on our local data stores. This hybrid solution results in cost savings, scalability, elasticity and flexibility, and presents data in a way that is most useful to researchers.

**What are the main challenges in promoting interoperability between different organisations, in order to homogenise responses in decision-making?**

Systems and tools should be open source and data available for research must be classified in a standardised way and properly safeguarded.

If data are well documented with high quality metadata and information about their provenance, context and quality (e.g. missing data and errors) comparability becomes more straightforward.

**What kind of efforts are you taking to make the knowledge gathered in the project more open, at both a national and international level?**

We have a number of tools we are using. These include a detailed blog and a series of formatted case studies from the projects to highlight the value of each project. We are also sharing the outcomes of the research through articles and presentations, as well as providing training on how we are able to scale up data curation practices. This project also benefits hugely from how effective and active the UK Data Service is in its communications, as well as the other Economic and Social Research Council (ERSC) and RCUK communication methods as we can tap into all these.

**What value will the project add to future energy sustainability and supply/demand decision-making worldwide?**

Household energy research undertaken at scale can help provide answers through linking powerful data sources to elucidate what is happening in fine-grained detail at the household level, for example, investigating consumer behaviour and patterns.



# Delving into data

*A new network led by the University of Essex and the University of Cape Town is building relationships between the UK and South Africa through investigating access to and use of household energy data for policy making and planning, ultimately promoting the long-term sustainable growth of South Africa*

Collaboration on research data infrastructure is increasingly becoming the norm as global partnerships between research institutes grow. One example of where this has delivered very successful outcomes is the partnership between the UK and South Africa on the knowledge exchange project 'Smarter Household Energy Data: Infrastructures for Policy and Planning'. This recently completed project resulted in collaboration on social science data services and the use of applied research to answer policy-relevant questions in the household energy domain.

Louise Corti, Associate Director at the UK Data Archive and Head of the Collections Development and Data Publishing teams for the UK Data Service (UKDS), was UK Principal Investigator working alongside Professor Martin Wittenberg, Director, DataFirst, University of Cape Town (UCT), who directed activities in South Africa. The team comprised Dr Nathan Cunningham lead on UKDS Big Data, Chris Park, UKDS Data Scientist, Simon Elam, Researcher at University College London, and others. Corti notes that this formal collaboration has allowed 'insight to be transferred bi-directionally and means that new research agendas are being stimulated by increasing the availability of energy data for social and economic research, as well as testing assumptions about the quality and international transferability of these data for research purposes.'

Funded by the Economic and Social Research Council (ESRC) in the UK and the National Research Foundation (NRF) in South Africa, the project started in November 2015 and has run for 18 months. The key centres in the partnership through this period were two social science data archives: the UK Data Archive at the University of Essex, curator

of the largest collection of digital data in the social sciences and humanities in the UK; and DataFirst at UCT, which archives and provides online research access to African microdata. 'Knowing more about domestic energy consumption can better inform strategies to deal with fuel demand and poverty, and strategies for energy savings and efficiency,' explains Corti.

## UNDERSTANDING ENERGY USE

Corti provides an insight into the team's work to study household energy consumption in rural South Africa and why it is important: 'Despite significant progress over the last 20 years, basic access to electricity in South Africa is by no means stable or guaranteed, and remains one of the largest development issues faced by post-apartheid South Africa.' The project was divided into a number of key deliverables. The first focused on Newton Fund Impacts, which promote economic development and social welfare of countries that receive official development assistance (ODA), that includes South Africa. The project aimed to strengthen data expertise and research partnerships between South Africa and the UK through the formation of a new network to investigate access to and use of household energy data for policy making and planning.

'The reasons for inequality in public service delivery, in rural as well as urban settings, particularly concerning access to electricity by poorer communities, are not as well understood as they could be. Knowing more about domestic energy consumption within households can better inform strategies to address fuel demand and poverty. For example, a richer, more policy-relevant picture can be gained by investigating the short-term dynamics of electricity access

using reconstructed complex large-scale longitudinal data linked to other sources of big data,' observes Corti.

The second deliverable explored knowledge exchange and researcher impact. This saw that the interdisciplinary nature of the project brought researchers together to learn from each other's methodological and analytic approaches. The third looked at policy-making impacts, namely through striving to identify and increase the amount of accessible and high-quality data which can help facilitate planning and policy making.

The fourth surrounded research data infrastructure and impacts for the research data profession, with the network bringing together UK and South African data professionals to share expertise and develop new research agendas. The goal was to strengthen innovation in the areas of exploiting big and complex data by providing joint activities between existing international centres of excellence. These enhance knowledge about how best to enable, acquire and provide access to new forms of data for social science and policy research.

The fifth deliverable developed capacity building impacts, through dedicated summer schools in both countries, which Corti says gave 'further impetus to the foundation of principles for the relatively young field of data science'. Community and engagement is the sixth deliverable, with the project 'benefitting from embedding within the existing infrastructures of the UK Data Service and DataFirst, both active in their use of up-to-date website content, blogs, social media outlets, mailing lists and collaborative

## LEARNING THROUGH SUMMER SCHOOLS

A key part of this project was the successful 'Encounters with Big Data' summer school courses. The core aim of the course was to bring together practitioners from a range of fields and train them in the use of big data technologies and applications. Two schools were run. These were held in February in Cape Town, South Africa in the UK's winter and in Essex, UK in the summer.

Corti believes these summer schools are an essential tool for delivering the outcomes for this project. 'The course introduced part of our work at the UKDS on 'scaling up' data curation and user access approaches for big data, predominantly where larger or more complex data sources, or where computationally intensive and iterative modelling is needed.'

The 'Encounters with Big Data' course covered aspects of extraction, exploration and statistical analysis of big data based on the approaches and technologies being used for the UKDS's new Data Services as a Platform (DSaaP) initiative, a Hadoop based data ecosystem. The new platform is using solutions that deliver data at scale, with speed and security, including Hive, Spark and Zeppelin, which integrate seamlessly with popular data analysis environments like R and Python.

communications with other stakeholders'. The seventh deliverable covered nurturing the collaboration, which was underpinned by leadership from the UK Data Service to ensure that the partnership developed and remained true to the joint vision.

## MULTIDISCIPLINARY WORK PACKAGES

In order to realise all of the key deliverables, the team divided the project into seven Work Packages (WPs). The first surrounded user needs and involves identifying key players and networks in the household energy research domain, consultation and expert workshops in the UK and South Africa and reporting on policy and researcher needs with collated real life use cases. WP2 involved scoping energy data. 'This included looking at availability and governance through consultation with data users, data owners and providers, a knowledge exchange workshop and reporting on data availability, licensing and access issues,' explains Corti.

*Knowing more about domestic energy consumption can better inform strategies to deal with fuel demand and poverty, energy saving and efficiency*

WP3 was about assessing data quality and entailed assessing the best tools for quality assessment, evaluating sample energy data, performing user testing of tools on quality assessment with researchers and data professionals and reporting on quality assessment. WP4 related to technical infrastructure. 'For this task we assessed data and user needs from a technical perspective, as well as testing data platforms and uploading samples of energy data,' notes Corti. WP5 involved pilot analysis projects and reported on analysis and limitations of data sources. WP6 concerned sustainability and compiled findings, and investigated future sustainable opportunities and pathways for securing timely access to quality energy data. Lastly, WP7 focused on training and capacity building.

## SUSTAINABLE GROWTH

The team hopes their work will have a far-reaching benefit, as Corti explains: 'Given the multidisciplinary nature of this project we are delivering a wide range of benefits that includes summer schools and workshops, blogs, published case studies and articles across the disciplines of the social sciences, data informatics and the natural environmental sciences.' Corti continues: 'For example, in liaison with our key stakeholders we hosted workshops to exchange methodologies for evaluating, curating and hosting household energy data, and to develop new skills for enabling, using and visualising these data.' Ultimately, the researchers hope that the work will help in the delivery of outcomes that promote the long-term sustainable growth of South Africa as an emerging knowledge economy.



## Project Insights

### FUNDING

Economic and Social Research Council

### CONTACT

Louise Corti

T: +44 1206872145

E: corti@essex.ac.uk

W: <http://gtr.rcuk.ac.uk:80/projects?ref=ES/No09886/1>

### BIO

Louise Corti is Associate Director at the UK Data Archive and leads the Collections Development and Data Publishing teams for the UK Data Service. She actively researches, publishes and edits regularly on key aspects of data management, sharing and reuse of social science data. Corti also acts as international consultant for new qualitative data archives setting up and is working to help the UK Data Service scale up for managing big data.

UK Data Service



DataFirst

