Facilitating access to Administrative and Big Data in the UK

The Administrative Data Research Network and the Big Data Network

Dr. Kakia Chatsiou (ADRN) & Dr. Aidan Condron (BDNS)
ESRC’s Big Data Network

Governance Structures

UK Data Service and Administrative Data Service

ADRC

Phase 1

Administrative Data

Data Research Centre

Data Research Centre

Data Research Centre

Data Research Centre

Third Sector Data Research Centre

Social Media Data Research Centre

International Centre for SM Research

ADRC

Phase 2

Business and Local Government Data

Data Research Centre

Data Research Centre

Third Sector and Social Media Data

Phase 3
Introducing the Administrative Data Research Network

Dr Kakia Chatsiou
User Services, ADS
The Administrative Data Research Network

- a UK-wide partnership between universities, government departments and agencies, national statistics authorities, the third sector, funders and researchers
- Facilitates access to linked de-identified administrative data for researchers

Dec 2011-2012 Administrative Data Taskforce Report

Oct 2013 Funding awarded

Nov 2013 - Nov 2014 Network process and policy development

Nov 2014 open for applications

Sept 2015 14 projects approved, 22 in progress
Eligibility criteria

Projects:

- non-commercial research purposes
- evident potential public benefit
- demonstrable value from using unit level administrative data to answer the research question
- not be more appropriately served by other existing services
- not normal operational functions undertaken by government departments or their agencies

People:

- academic, govt researcher 3rd sector
- accredited/safe
How It Works

Safely
Lawfully
Ethically
Securely
Challenges

- Development of procedures, new service – opened Nov 2014
- Data discovery
- Final decision on access rests with the data custodian
  - Dependent on legal gateway, resources and risk appetite of data custodian
- Time scales
Get in touch

- Learn more about us and our work with admin data at RSS/ADRN joint event 08/10/2015, London: https://adrn.ac.uk/news-events/latest-news/adrn-rssevent

- Contact ADRN User Services: help@adrn.ac.uk
- Follow us: @adrn_uk
- Sign up for our newsletter http://adrn.ac.uk
UK Data Service
Big Data Network Support (BDNS)

Aidan Condron
Collections Development and Producer Relations,
Big Data Network Support
UK Data Archive
University of Essex
BDN2 Goal

‘The overarching aim of the ESRC Big Data Network is to build infrastructure which enables researchers to exploit the richness of the big data landscape in the UK to advance research in the social sciences and economics and to enhance its societal impact’

DISAC 03/14, November 2014
Process Harmonisation

Data storage, discovery and access, analysis and visualisation, user training
Open Source Data Platform

HDP 2.3
Hortonworks Data Platform

GOVERNANCE & INTEGRATION
- Data Lifecycle & Governance
  - Falcon
  - Atlas
- Data Workflow
  - Sqoop
  - Flume
  - Kafka
  - NFS
  - WebHDFS

DATA ACCESS
- Batch
  - MapReduce
- Script
  - Pig
- SQL
  - Hive
- NoSQL
  - HBase
  - Accumulo
  - Phoenix
- Stream
  - Storm
- Search
  - Solr
- In-Memory
  - Spark
- Others
  - ISV engines

SECURITY
- Administration
- Authentication
- Authorization
- Audit
- Data Protection
  - Ranger
  - Knox
  - Atlas
  - HDFS TDE

OPERATIONS
- Provision, Manage & Monitor
  - Ambari
  - Cloudbreak
  - Zookeeper
- Scheduling
  - Oozie

YARN: Data Operating System

HDFS
(Hadoop Distributed File System)

DATA MANAGEMENT
- Linux
- Windows
- Deployment Choice
- On-Premise
- Cloud

UK Data Service
Further Contact

• Data Curation and Technical
  Ben Newman-Wright – bwrighp@essex.ac.uk
  01206 874254

• Training
  Sarah King-Hele – sarah.king-hele@manchester.ac.uk
  0161 275 0279

• Engagement:
  Aidan Condron – acondron@essex.ac.uk
  01206 874254