

How will open data impact on records management?

Professor Elizabeth Shepherd
Department of Information Studies, UCL
e.shepherd@ucl.ac.uk

UK administrative data context

- Freedom of Information Act 2000 (FOIA) fully in force 2005 for all public authorities
- Data Protection Act 1998
- Public Records Act 1967 records closure for 30 years, being reduced to 20 years
- Open government data movement
- OPSI and TNA
- ICO governance



UCL DIS research projects

- ADRC-E: Information governance for social and economic research using government administrative data
- ITrust: Case Study: The Role of the Records Manager in an Open Government Environment in the UK

ADRC-E

Vision

To benefit society and quality of life through intelligent use of administrative data



Administrative Data
Research Centre
England

Administrative Data Research Centre for England

UNIVERSITY OF
Southampton



A bar chart icon with four vertical bars of increasing height from left to right, colored in shades of green.
Institute for
Fiscal Studies

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



Leading education
and social research
Institute of Education
University of London



People at ADRC-E include:

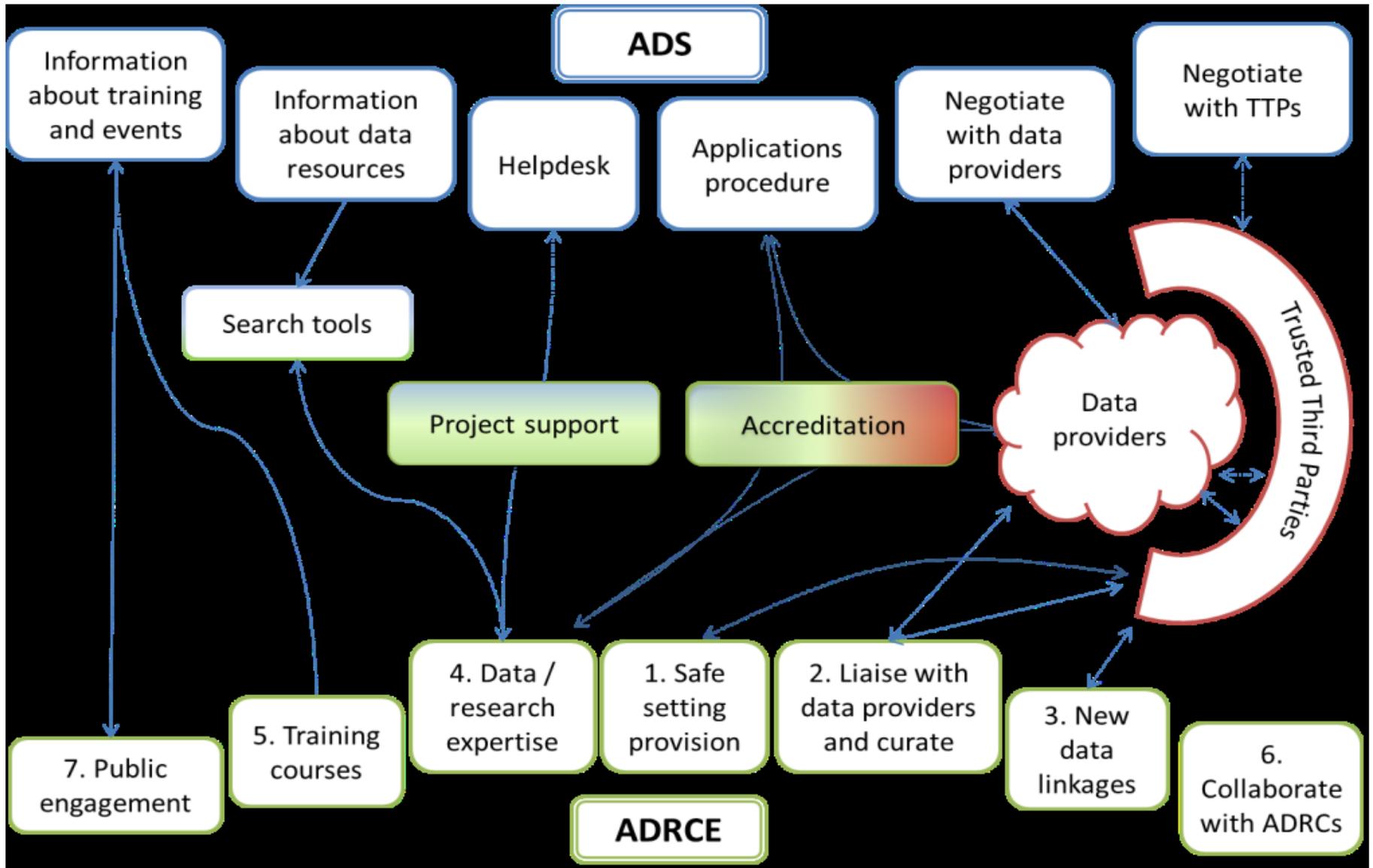
- Director: Peter Smith (Univ of Southampton)
- Deputy Directors: Ruth Gilbert (UCL), Dave Martin (UoS), Ben Humberstone (ONS)
- Associate Director – Operations: Emma White (UoS)
- Management Group: Lorraine Dearden (IoE, IFS), Paul Wilkinson (LSHTM), Spiros Denaxas (UCL), Gabi Durrant (UoS), Oz Parchment (UoS)
- Co-investigators, Data Scientists and Research Fellows

Administrative Data Research Network

- ADRN Governing Board
- ADS – UKDS (Essex, Manchester), ADLS (St Andrew/Edinburgh, Manchester and Oxford)
- ADRC Northern Ireland – Queen’s Belfast, University of Ulster, Northern Ireland Statistical Research Agency
- ADRC Scotland – Aberdeen, Dundee, Edinburgh, Glasgow Herriot-Watt, St Andrews, Scottish Government
- ADRC Wales – Swansea, Cardiff, Bristol, Welsh Government

Five integrated work-streams of ADRC-E

1. Facilities including e-Infrastructure
2. Data resources and management and statistical support
3. Research (eg data linkage; linked data quality; **IG**)
4. Training, capacity building
5. Stakeholder/public engagement; collaboration with ADRN



Five year vision and beyond

Established & internationally recognised centre for linked admin data research in UK network

Excellent & rewarding service for research users

Trusted & well-understood partner for data providers

Range of well-developed protocols for handling projects of differing data type, complexity and sensitivity

Acknowledged centre of linked data methodological expertise

Research question: what are the requirements for information governance which ADRC-E should reflect in its national e-infrastructure policies and practices?

- Aim: To develop a good practice framework for curating linked data for ADRC-E which helps to establish information governance policy landscape; best policy practice; and reflects community and user attitudes to government data, with a focus on the widening use agenda.



Administrative Data
Research Centre
England

Data, data, everywhere

Various overlapping terms to describe data:

- Administrative data
- Big data
- Open data
- Linked data
- Official Statistics
- Public Records

Data, data, everywhere

Data characteristics:

- Personal
- Sensitive
- Confidential
- Classified
- Safeguarded
- Disclosive
- Controlled
- Anonymised
- De-identified
- Pseudonymised
- Safe
- Secure
- Open
- Linked

Administrative data

"Administrative data refers to information collected primarily for administrative (not research) purposes... by government departments and other organisations for the purposes of registration, transaction and record keeping, usually during the delivery of a service." (www.adls.ac.uk)

“data ...containing information which is not primarily collected for statistical purposes”

(United Nations Economic Commission for Europe (UNECE) ASSIST Handbook)

Big data; open data

Big data

"Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization" (Gartner)

Open data

"Open data is data that can be freely used, shared and built-on by anyone, anywhere, for any purpose." (Open Knowledge Foundation)

Linked data

“Data linkage is the joining of two or more administrative or survey datasets using individual reference numbers/identifiers or statistical methods such as probabilistic matching” (Administrative Data Taskforce 2012)

Most Linked Data are open, but they don't have to be

Most Open Data aren't Linked

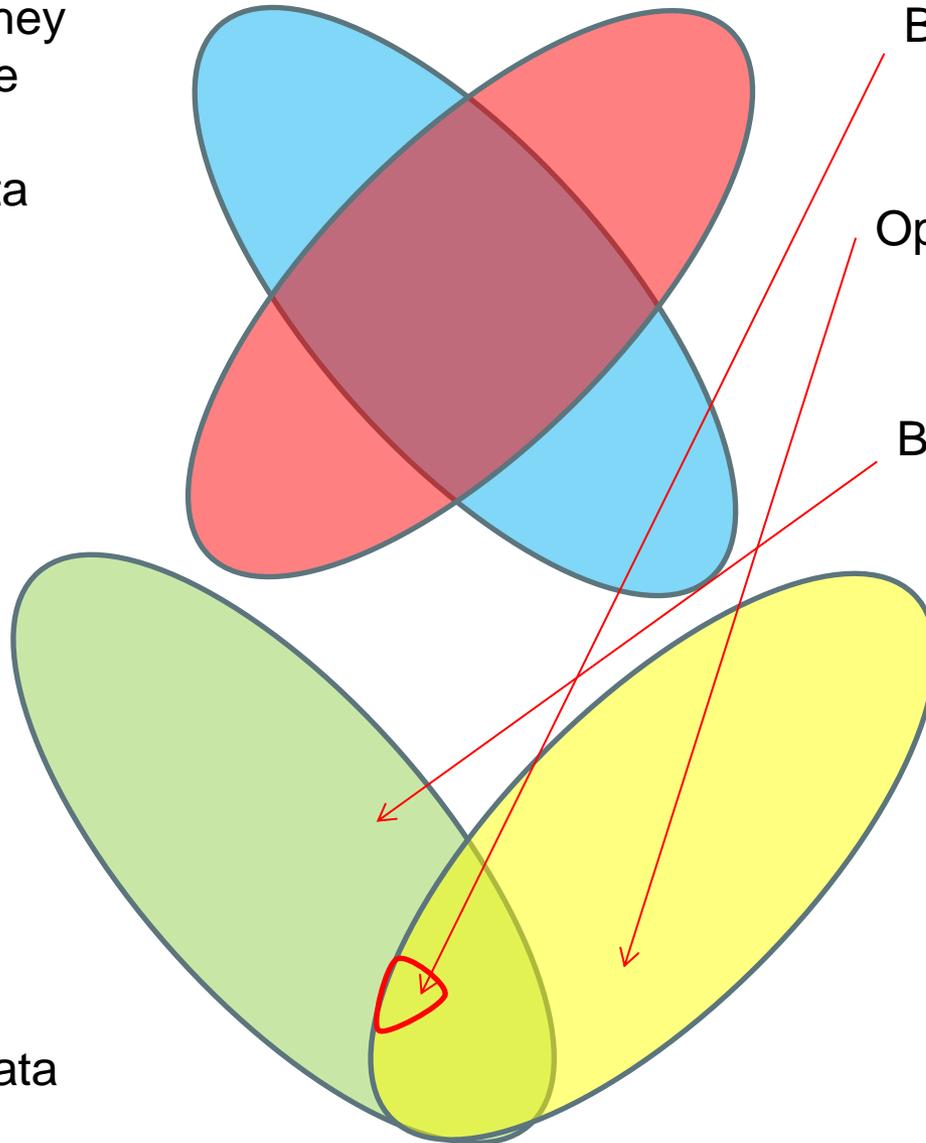
Some Admin data are Big

OPEN

LINKED

BIG

ADMIN



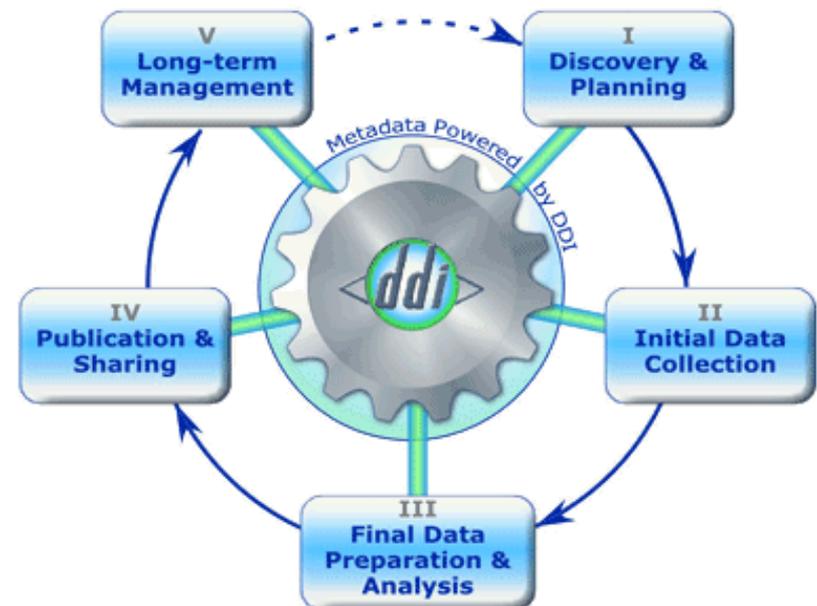
Big Linked Admin, not Open

Open Admin, not Big or Linked

Big Open Linked, not Admin

ADRC-E research themes

- A risk management approach to disclosure review
- With/out consent
- Trusting government administrative data



What challenges been identified so far?

- **Data consistency & reliability over time**

“some data, and this is particularly a concern with administrative data, that administrative data aren’t archived very well. ...they get updated, they get modified, ...someone coming back and saying ..., I want the same set of data for the same period, it might be the case that there are differences... between what that data now looks like and what the data looked like [before]”

Access to datasets: more or less restricted?

“before Fols you could email local authority ... teams that were responsible for an area that you were interested in, have a conversation, say I’m doing this or that research and I’d like this, and they were often quite friendly and helping you. Now, you know, there’s no contact there, it’s all done through the FOL team and its just sort of very impersonal.’

[A3]

Access to datasets: more or less restricted?

“increased government sensitivity about data and I think better data ...protocols in government. But this was late 90s, and government agencies didn't take data security as seriously as they do now.” [A1]

“culture of fear around privacy” [A3]

Access to datasets: more or less restricted?

“there used to be a very small pipe and you know there were certain things you could get, and they were in quite archaic but well known places. Now we’ve got a very big pipe that sort of draws things together but sometimes you don’t realise quite what’s out there because, you know, the sort of documentation is not quite in place or you know the ways of finding it are not quite there.” [A3]

What is InterPARES Trust?

- multi-national, interdisciplinary research project led by UBC, Canada, funded by SSHRC
- exploring issues around digital records entrusted to the Internet
- research partnership of fifty universities and organizations, public and private, in North America, Latin America, Europe, Africa, Australasia & Asia

<http://www.interparestrust.org/>





Case Study: The Role of the Records Manager in an Open Government Environment in the UK

- To establish a picture of existing practice in information and records management in local government
- To examine the recordkeeper's role in open government data
- To support better understanding of the recordkeeping issues and records policies which take account of open government requirements

Lack of clarity about the governing policy: *Local Government Transparency Code (2014)*

- Useful for determining in general which data need to be published: **but**
- Code assumes established electronic systems, whereas many RK systems are primarily paper-based
- Code emphasizes publishing as much data as possible, rather than data which will be utilised
- Lack of guidance on open government **data**
- Lack of sector specific guidance

Who is responsible for open government data in the public authority?



What are the main challenges of the open government data environment?

- Risk management and data security
- Protecting personal data
- Cleaning up data before proactive release
- Ensuring future access to open data
- Lack of consistent policy on release by different teams
- Data accuracy
- Metadata standards for open government data

What does this mean for the records management profession?

- An important records management role, but not traditional paper file management
- Job titles and job descriptions: convergence with other information professions or distinctive?
- What kinds of skills and knowledge are needed?
- How to raise those skills & knowledge in the records profession?

How will open data impact on records management?

Professor Elizabeth Shepherd
Department of Information Studies, UCL
e.shepherd@ucl.ac.uk