A tale of research data management in practice

Veerle Van den Eynden
UK Data Archive
University of Essex

Embracing Data Management - Bridging the Gap Between Theory and Practice
VLIR-FOSTER seminar
Brussels, 4 June 2015

UK Data Service

FOSTER
FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH

Science advances through data sharing

ELSEVIER

Energy Policy

Volume 62, November 2013, Pages 1204–1211

Public perceptions of climate change and energy futures before and after the Fukushima accident: A comparison between Britain and Japan

Wouter Poortinga[a, b], Midori Aoyagi[c], Nick F. Pidgeon[b]

highlights

- We report data from 2005 to 2011 of British and Japanese attitudes towards nuclear power and climate change.
- The Japanese are less supportive of nuclear power as a solution to climate change than the British.
- Public support for and trust in nuclear power has collapsed in Japan after Fukushima.
- British public attitudes to nuclear power are remarkably robust in the wake of Fukushima.

Data used (national surveys): Public Risk Perceptions, Climate Change and the Reframing of UK Energy Policy in Britain, 2005; Public Perceptions of Climate Change and Energy Futures in Britain, 2010; …

Society benefits from data sharing

F1000Research
Open for Science

F1000.com

Articles | Collections | For Authors | For Referees | About | Blog

What is open science? | New on F1000Research – 17 November 2014

Beat Ebola with better research sharing, says discoverer of virus

POSTED BY THOMAS INGRAHAM, 12 NOVEMBER 2014

COMMENTS 0 | Share 418

Search Posts

Follow F1000Research

Key points

- *F1000Research: Ebola* article collection launched to enable all Ebola-related knowledge to be made available on a single platform within days of submission. All articles submitted to this collection will be prioritized and those accepted will be published free of charge.
- *F1000Research* has set up a dedicated phone line (+442071931030) to enable those working in affected areas to dictate their reports directly to our office.
- F1000Prime makes all existing and upcoming Ebola and Marburg virus Article Recommendations free to access.

Sharing critical information faster

Peter Piot (Director the London School of Hygiene & Tropical Medicine and co-discoverer of the Ebola virus) has published an Editorial in *F1000Research* appealing for everyone working on the current epidemic to publish all their findings and experiences rapidly and openly.

Since the onset of this year's West African outbreak* there have been over 13,250 suspected cases and 4,960 deaths, the majority of these in the last two months. Depending on how effective control efforts are over the next few weeks the number of cases will likely swell to between 16,500 and 30,000 by December.

Speed is required at all levels if this epidemic is to be successfully controlled; identification and isolation of cases, clinic construction, and community action campaigns must all be carried out rapidly and on a large scale. Speed is just as vital to Ebola research; we need knowledge to spread faster than the virus if Ebola's impact is to be mitigated. For this to happen, the whole research cycle has to start spinning faster.

Some major funders have responded quickly by releasing emergency research grants, such as the NSF's RAPID awards and the Wellcome Trust's Emergency Ebola Initiative. Science publishing needs to be just as nimble, but in the current system it can take several months for findings to be made publically available; with Ebola cases doubling every 20 days or so, such delays are unacceptable.

BBC | Sign in | News | Sport | Weather | iPlayer | TV | Rad

NEWS

Home | UK | World | Business | Politics | Tech | Science | Health | Education | Entertai

Asia | China | India

How 'crisis mapping' is helping relief efforts in Nepal

By Saira Asher
BBC News

6 May 2015 | Asia

Kathmandu, and nearby areas before and after mapping efforts were ramped up after the quake

Thousands of people in remote parts of Nepal are still in need of medical help and basic supplies. But with roads damaged and buildings collapsed, knowing what aid is needed and where, is a challenge. One group of Nepalis, backed by a global community, is trying to change that by "crisis mapping" Nepal

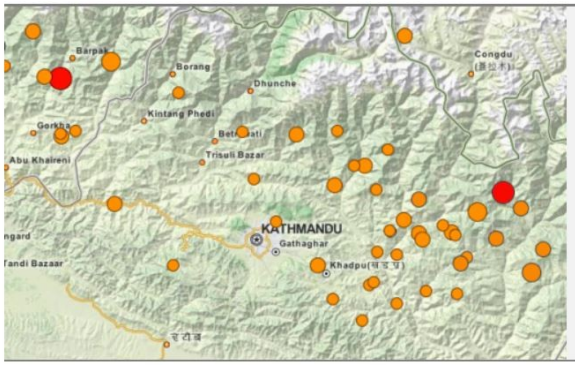
Nepal quake

Unsettled Earth continues to rattle Nepal

Nepal quake causes

HOT Humanitarian Open Street Map Team

About | Update



FEATURED PROJECT

Nepal 2015 Earthquake Response

The **2015 Nepal earthquake** (Wikipedia) struck on 25th April with a magnitude of 7.8, followed be many aftershocks including a large magnitude 7.3 quake on 12 May. The quake hit the city of Kathmandu and the mountainous regions to the north.

The Humanitarian OpenStreetMap Team [HOT] applies the principles of open source and open data sharing for humanitarian response and economic development.

UK Data Service

Helping researchers manage and share data

- Data management = organisation, documentation, storage, safeguarding, preservation and accessibility of data, incl. ethical and legal aspects of data handling and data ownership

- Data sharing = release of data for use by other people

Why manage research data well ?

- Data creation in research is often expensive
- Data = cornerstone of research
- Data underpin published findings
- Good quality data = good quality research
- Protect data from loss, destruction,…
- Compliance with ethical codes, data protection laws, journal requirements, funder policies

Research integrity

Openness

UK Data Service

Boost factors: research funders, EU

European open access policies: Horizon 2020, European Research Council (ERC)

- communication & recommendation on access to / preservation of scientific information (July 2012) (publications & research data)
- pilot on open access to research data, primarily data underlying (open access) scientific publications for Horizon 2020
- data management guidelines for Horizon 2020 (~ policies)

generally based on

OECD Principles and Guidelines for Access to Research Data from Public Funding

UK Data Service

Boost factors: research funders, UK

- *Publicly funded research data are a public good, produced in the public interest, that should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.*
- in accordance with relevant standards and community best practice
- metadata to make research data discoverable
- legal, ethical, commercial constraints on release of research data
- recognition for collecting & analysing data; limited privileged use
- acknowledge sources of data, intellectual contributions, terms & conditions
- use public funds to support the management and sharing of publicly-funded research data

Research Councils UK Common Principles on Data Policy (May 2011)

Boost factors: data infrastructure, UK

Research funders invest in data support services and infrastructure, e.g.

- UK Data Service (ESRC)
- NERC data centres (NERC)
- MRC Data Support Service
- Genbank (BBSRC, MRC)
- Atlas Petabyte Storage (STFC)
- Archaeology Data Service (AHRC)

UK data centres

- Archaeology Data Service
- Biomedical Informatics Research Network Data Repository
- British Atmospheric Data Centre
- British Library National Sound Archive
- British Oceanographic Data Centre
- Cambridge Crystallographic Data Centre
- ChemSpider
- ChemSpider Synthetic Pages
- eCrystals
- Endangered Language Archive
- Environmental Information Data Centre
- Ethno-ornithology World Archive
- National Biodiversity Network
- National Geoscience Data Centre
- NERC Earth Observation Data Centre
- NERC Environmental Bioinformatics Centre
- Polar Data Centre
- The Oxford Text Archive
- UK Data Archive
- UK Solar System Data Centre
- Visual Arts Data Service

Boost factors: training

- Research Data Management Training MANTRA (Edinburgh) – online learning units
  - http://datalib.edina.ac.uk/mantra/

- Digital Curation Centre:
  - Data management planning http://www.dcc.ac.uk/resources/data-management-plans
  - DMP Online tool: https://dmponline.dcc.ac.uk/
  - Data management training / courses http://www.dcc.ac.uk/training/data-management-courses-and-training

UK Data Service

- Curator of the largest collection of digital data in the social sciences and humanities in the UK
- UK Data Archive (www.data-archive.ac.uk) lead organisation in a network
- Based at University of Essex, essentially as department

- Extensive experience of supporting researchers and other creators of social science data (and related disciplines)
- We manage data sharing for the ESRC (since 1995)

- Our best practice approaches to making data shareable are based on:
  - challenges faced by researchers to share data
  - archiving research data – quantitative and qualitative

www.ukdataservice.ac.uk

UK Data Service

What we do in practice

Research & strategy:

- ESRC research data policy co-development
- Research on data sharing practices

e.g. Incentives and motivations for sharing research data: a researcher's perspective (http://t.co/K6P006cROH)

Analyse needs

- evaluate existing data management practices
- engage with and work with researchers
- identify obstacles to data sharing

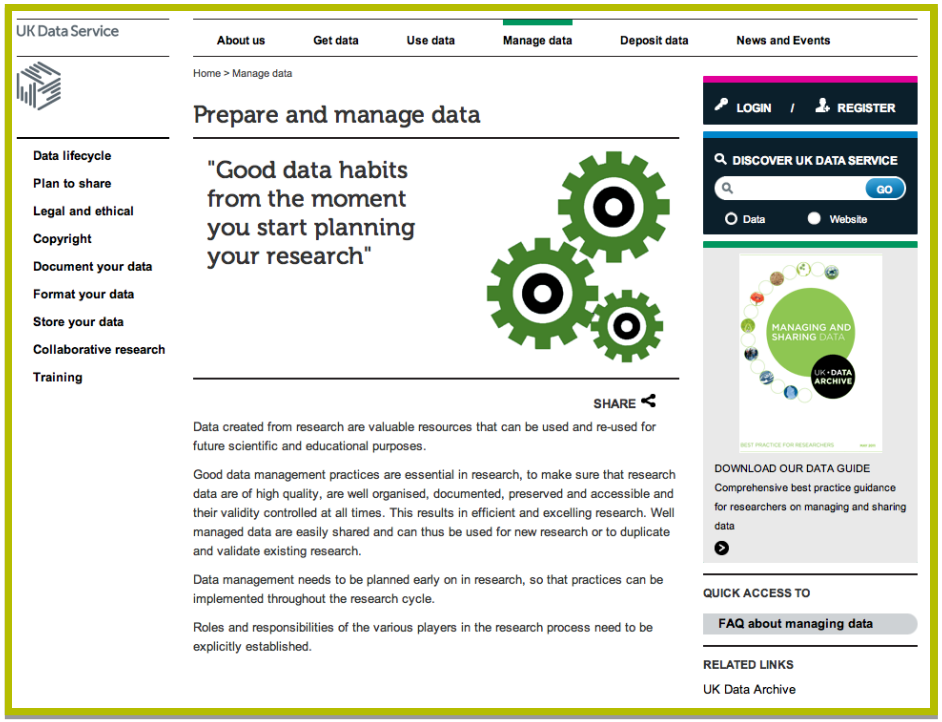
Provide solutions for researchers & institutions

- practical strategies to embed data management into research practices
- tools and templates
- guidance, training and bespoke advice
- help control archive ingest costs (largest share of costs, more than access and preservation)

UK Data Service

Our data management guidance

- Online best practice guidance: ukdataservice.ac.uk/manage-data.aspx
- Managing and Sharing Research Data – a Guide to Good Practice: (Sage Publications Ltd)
- Helpdesk for queries: ukdataservice.ac.uk/help/get-in-touch.aspx
- Training: www.data-archive.ac.uk/create-manage/advice-training/events



UK Data Service

Our guidance

- plan to share research data
- legal and ethical aspects of data sharing and reuse
- data copyright
- documentation and metadata to understand and use data
- file formats, organising, versioning and quality control
- storage, backup, encryption and security of data and files
- strategies for collaborative research

A taster of our guidance

Document your data

"Make data clear to understand and easy to use"

SHARE

Overview

A crucial part of making data user-friendly, shareable and with long-lasting usability is to ensure they can be understood and interpreted by any user. This requires clear data description, annotation, contextual information and documentation.

Study-level documentation

Study-level documentation provides an overview of the research context and design, data collection methods, data preparation and results or findings and is key to enabling the secondary user to make informed use of the data.

Data-level documentation

Labelling and documenting individual data items - numerical, textual or audio-visual - can often be embedded within a data collection or recorded in an accompanying document.

Options for sharing confidential data

- Obtain informed consent, also for data sharing and preservation / curation
- Protect identities e.g. anonymisation, not collecting personal data
- Regulate access where needed (all or part of data) e.g. by group, use, time period
- Securely store personal or sensitive data (separately)

Consent needed across the data life cycle

- Engagement in the research process
  - decide who approves final versions of transcripts
- Dissemination in presentations, publications, the web
  - decide who approves research outputs
- Data sharing and archiving
  - consider future uses of data

Always dependent on the research context – special cases for covert research, verbal consent, etc.

Anonymising data

- Direct identifiers – often not essential research info
- Indirect identifiers

- Remove / pseudonymise direct identifiers

  e.g. names, address, institution, photo

  - reduce the precision/detail of a variable through aggregation

  e.g. birth year vs. date of birth, occupational categories, area rather than village

  - generalise meaning of detailed text variable

  e.g. occupational expertise

  - restrict upper lower ranges of a variable to hide outliers

  e.g. income, age

Managing data access

- UK Data Service: web access to data and metadata
- Data freely available for use; commercial use charges
- Metadata / documentation always open
- Data available under 3 access levels:

  OPEN

  SAFEGUARDED – End User Licence

  (e.g. not identify any potentially identifiable individuals)

  - Special agreements: depositor permission; approved researcher
  - Embargo for fixed time period

  CONTROLLED – only for accredited users

  - Access via on-site or virtual secure environment

  (secure lab

ESRC data management plan

- Assessment of existing data
- Information on new data
- Quality assurance of data
- Backup and security of data
- Difficulties in data sharing and measures to overcome these
- Consent, anonymisation, re-use strategies
- Copyright / Intellectual Property Ownership
- Responsibilities
- Management and curation

Data management planning for ESRC researchers

Data management plan | ESRC data policy | ESRC research centres

SHARE

All ESRC grant applicants generating data during their research have to include a data management plan with their Je-S grant application.

A data management plan helps to decide how research data will be managed throughout the research cycle and will be available for sharing afterwards. Most research data can be successfully archived and shared.

ESRC expects grant holders to consider all issues related to confidentiality, ethics, security and copyright before initiating the research. Any challenges to data sharing (e.g. copyright or data confidentiality) should be critically considered in a plan, with possible solutions discussed to optimise data sharing.

A data management plan includes the following topics.

ASSESSMENT OF EXISTING DATA

- an explanation of the existing data sources that will be used by the research project, with references
- an analysis of the gaps identified between the currently available and required data for the research

Where research grant applicants plan to create new data as part of their ESRC-funded proposal, they must demonstrate that no suitable data are available for re-use. ESRC encourages the re-use of existing data and therefore encourages applicants and grant holders to consider the breadth of data available from various sources before committing to primary data collection.

When using existing data sources, consider copyright and IPR of those data and the conditions of their use, to decide whether resulting derived data can be shared.

Data sources that can be consulted are:

- Discover UK Data Service, with over 6,000 datasets of key economic, social and historical data spanning many disciplines and themes
- RCUK Gateway to Research of past and present research grants and their outputs

INFORMATION ON NEW DATA

Provide information on the data that will be produced or accessed by the research project, e.g.

- data volume
- data type
- data quality, formats, standards documentation and metadata
- methodologies for data collection and/or processing
- source and trustworthiness of third party data

Using standardised and interchangeable data formats ensures the long-term usability of data. Clear and detailed data descriptions and annotation, together with user-friendly accompanying documentation on methods and contextual information, makes data easy to understand and interpret and therefore shareable and with long-lasting usability.

Guidance on data formats

Guidance on documenting data

UK Data Service

Tools & templates

- Model consent form: http://www.data-archive.ac.uk/media/112638/ukdamodelconsent.pdf
- Survey consent statement: http://data-archive.ac.uk/media/147338/ukdasurveyconsent.doc
- Transcription template: http://data-archive.ac.uk/media/136055/ukdamodeltranscript.pdf
- Transcription instructions: http://data-archive.ac.uk/media/285633/ukda-example-transcription-instructions.pdf
- Transcription confidentiality agreement: http://data-archive.ac.uk/media/285636/ukda-transcriber-confidentiality-agreement.pdf

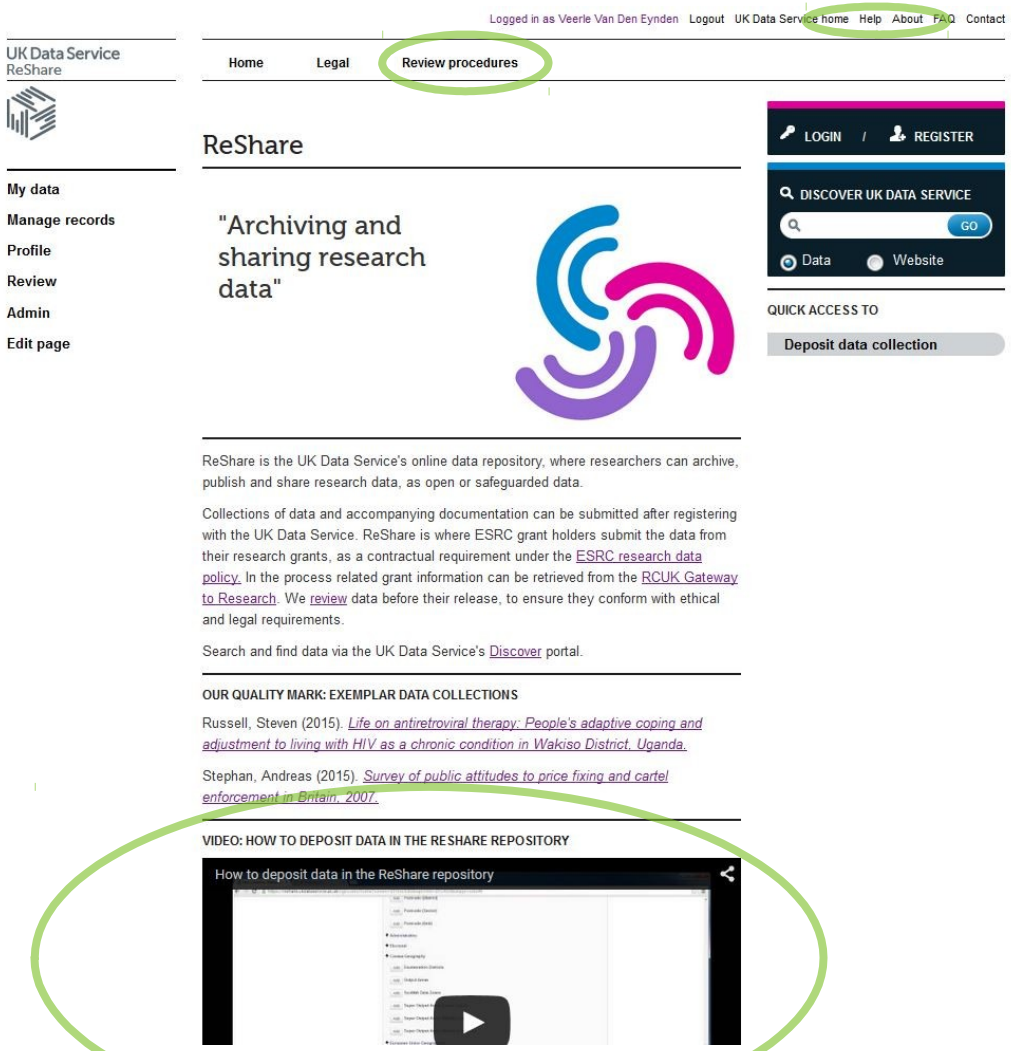
- Data list template: http://data-archive.ac.uk/media/2989/UK%20Data%20Archive%20Example%20Data%20List.pdf

- RDM costing tool: www.data-archive.ac.uk/media/247429/costingtool.pdf

Data management training

- Regular workshops on ‘Managing and Sharing Research Data’
- Bespoke training events by invitation
  - ICPSR summer school Curating and Managing Research Data for Reuse (July 2015): http://www.icpsr.umich.edu/icpsrweb/sumprog/courses/0149
  - Doctoral training: managing and sharing research data, University of Ghent, Faculty of Psychology and Educational Sciences (Dec 2014): http://ukdataservice.ac.uk/news-and-events/eventsitem/?id=3914
  - FOSTER-CESSDA RDM doctoral training series, Lausanne, Ljubljana, Cologne, Manchester (May-Nov 2015): http://ukdataservice.ac.uk/about-us/projects/foster-cessda-training/details.aspx

ReShare – help guidance



UK Data Service

Questions

Contact details

veerle@essex.ac.uk