Unleashing Data at RepoFringe 2016

The Unleashing Data World Café workshop took place at Repository Fringe in Edinburgh on the 2\textsuperscript{nd} of August 2016. Participants discussed ten questions on the topic of how research data could be used to achieve greater impact.

Contents

1. How can repositories help researchers achieve impact with their data? ........................................1
2. What are the strengths of the repository community that help us support researchers to achieve impact from data / making a difference with their data? ....................................................2
3. What are the weaknesses / barriers that hold the repository community back from helping researchers to make a difference with data? .............................................................................2
4. What are the opportunities for the repository community to make a difference with data? 2
5. What are the risks or costs of trying to facilitate others to make a difference with data? .... 2
6. Working with business to make a difference with data. .................................................................3
7. What role can research funders play to help repositories help researchers make a difference with data? .......................................................................................................................................3
8. What role can universities play to help repositories help researchers make a difference with data? ..............................................................................................................................................4
9. How can the repository community use crowd-sourcing (e.g. citizen science) to engage the public in re-use of data? ..............................................................................................................4
10. Which is the best example of unleashing data in the past 12 months? ........................................5

1. How can repositories help researchers achieve impact with their data?
(discussion led by Prof Andrew Millar)

- Non-publication route to ‘open’.
- Trust.
- Service-oriented. But need researcher too.
- Contact all stages of research*: ECR and admin and PIs.
  * Teaching/OER – use data to learn.
- Impact depends on disciplinary insight among staff managing repository (decline of subject librarian).
- Subject librarians as colleagues.
- Advise on route e.g. lay summary.
- Keep access statistics.
- Bench\(\ing\) FAIRdom.
2. **What are the strengths of the repository community that help us support researchers to achieve impact from data / making a difference with their data?**
   - Bench\ing FAIRdom.

3. **What are the weaknesses / barriers that hold the repository community back from helping researchers to make a difference with data?**
   - Ambiguity over who manages the data (researchers / IT / Librarians) can cause inertia.
   - Possessive attitude of researchers to their data, can be worsened by the digital context, and leads to a low profile in some cases.
   - Intellectual Property Rights/Copyright.
   - Poor management of data by researchers. Cultural change is needed to tackle this problem.
   - Resourcing gap (especially technical resources including for development of software).
   - Commercial market monopolisation. Including lack of configurability / customisability of products.
   - Limited engagement from researchers. Need resources to support advocacy and outreach.
   - Compliance culture.
   - Lack of credible sanctions for ‘poor’ practice.
   - Follow-up when datasets are deposited.
   - Lack of dialogue between researchers and RDM teams.

4. **What are the opportunities for the repository community to make a difference with data?**
   - Employ standards and agreed terminology (re discovery, access, re-use and long-term preservation and re-use).
   - Encouraging re-use / exploitation (and curation) both internally and externally.
   - Raising awareness of the potential of data to academics.
   - Lessons to be learned from Open Access repositories.
   - Supporting / promoting use of repositories for digital content more generally – reduce the emphasis on publications.
   - Examples / storytelling of re-use through press office, PR, twitter.
   - Feeding back to academics on re-use using data metrics.
   - Monitor citation of datasets (DOIs).
   - Altmetrics on datasets.
   - Serving a real need to certain data/research communities.

5. **What are the risks or costs of trying to facilitate others to make a difference with data?**
   - Sensitive data:
     - Need to obtain consent of subject;
     - Need data controller;
     - Risk of data combination leading to de-identification of subjects;
     - Cost of controlled access.
• The person who pays is not the beneficiary with respect to:
  o Deposit
  o Retention
  o Curation
  o Preservation
  o Access

• Need researcher education on:
  o Benefits
  o Recognition
  o Culture change
  o Alternative use.

6. Working with business to make a difference with data.
   (discussion led by David Richardson)

   To follow.

7. What role can research funders play to help repositories help researchers make a
difference with data?
   • ESRC mode of providing repository.
   • Clearer guidance e.g. what is compliant? (e.g. Supplementary data)
   • More discipline-specific repositories e.g. BGS [British Geological Survey], UKDA [i.e. the ESRC’s UK Data Archive] etc.
   • Provision of data selection criteria i.e. following NERC’s example as good practice.
   • Follow up with researchers around data deposits e.g. a form of peer review.
   • EPSRC mandates [researchers to provide] a Data Management Plan and provide their own template and guidance.
     o The same requirements for all funders.
   • Examples from funders e.g. case studies demonstrating impact.
   • Funders could harmonise and join up requirements so as to make it clearer to services and researchers what the funders want.
   • Cost / funding
     o Clarifying what costs are available and through which channels.
     o Highlight and give expected costs for storage etc.
     o Support for long-term preservation.
   • Funders are creating problems for the future: “silod data”. Risk of loss of the research outputs from collaborative projects.
   • Better phrases [?]:
     o Fragmentation
     o Collaboration breakdown
     o End of project which does not meet the RCUK common principals fully.
   • Enable data to be available closer to the point of creation (i.e. intelligent, not arbitrary, embargo).
   • ESRC Costing Model adoption.
   • Auditing of compliance with policy.
8. What role can universities play to help repositories help researchers make a difference with data?

- Strong guidance on which repositories are appropriate for each dataset.
- Good metadata for easy discoverability and availability.
- Encourage data sharing of student projects.
- Offer ‘carrots’.
- Provide case studies, examples of impact.
- Promote what data is held.
- Encouraging students to store datasets in the appropriate repositories.
- Embed good practice.
- Identifying champions: “Data Impact Ambassadors” to share their stories.
- Provide training on impact: how it is measured for REF, and good practice in achieving impact.
- Foster better links with public relations / media team.
- Put funding into the repository team to work on impact.
- Provide resources to incorporate or develop repository functionality that can support impact such as Altmetrics or visualisation tools.
- Recognition of data sharing as part of researchers’ professional academic development, to reward e.g. visibility of datasets and numbers of data citations.
- Universities could work with funders e.g. with respect to REF.
- Development of strategy and policy.
- Engage with undergraduates (as researchers) as well as other researchers.

9. How can the repository community use crowd-sourcing (e.g. citizen science) to engage the public in re-use of data?

(discussion led by Nicola Osborne)

- EQUAL TOP IDEA: Data Café / f2f (face-to-face) re-use, sociable re-use of data.
- EQUAL TOP IDEA: Linking toolkit / blocks of data e.g. API-driven IFTTT / Scratch interface.
• THIRD TOP IDEA: Incorporate data re-use and audience in RDM Plans [sic researcher’s own Data Management Plan?].
• Normalisation of data and/or APIs for all (all data is different). IFTTT for data / MS flow. Dev blocks. Also scratch style toolkit for research data across multiple repositories e.g. (Edinburgh) CityScope etc.
• Advocates for change / Open Data: publishers and researchers.
• Link to other existing resources e.g. research group / Lab web pages.
• Crowdsourced ontologies.
• Timeline and sustainability challenge.
• Who pays / business case / value of impact.
• Funding via crowdfunding? For useful re-use of data.
• Gamification.
• Generating ideas: crowd voting / ideascale.
• New additions: Why of benefit. Re-use for what?
• Benefit / reward / motivation.
• Funding mechanism: or What would you pay?
• Physical roadshows i.e. tangible e.g. Eisteddfod / Science Fairs / Makers’ Faires.
• Data exhibits / installations.
• Maker Faire / similar formats / hack / hackathon.
• Who should organise such events: repositories / academic / university?
• Recognition of re-use: for community and originator (motivating and rewarding)
  o Bring wacky data use
  o Awards for re-use at RepoFringe 2017
• Audience flag: Educate researchers and describe data better for re-use.
• Not just custodians but promote that data, seek opportunities or calls etc.
• Focus on ongoing work and data?
• Include re-use in RDM plans: ideas, updates, flag in repository.
• Traditional citizen science: link e.g. PURE to other impacts / use.
• Traditional citizen science (and scale data gathering such as twitter) gathers data. What about re-use?! What next?
• Book crossing for data? (inc DOI?) (question of authenticity and versioning). Maybe use QR code for DOI and location, for measure of impact, must be eye-catching.
• AR (Augmented Reality e.g. Pokémon Go) for data discovery.
• Challenge: most interesting data not always accessible.
• Expectation management.

10. Which is the best example of unleashing data in the past 12 months?
• TOP ANSWER: The 1641 Depositions. (slightly more than 12 months ago). Have generated extensive public interest and debate. The web presence includes a specialised search allowing users to search for individuals by name.
• Star mapping
• CRUK ‘Spot the tumour’
• Need to take account of lead-in times
- What’s ‘best’?
- Best packaging?(!)
- The Referendum(!)
- The Panama Papers(!)