Report out to ROI subgroup on JISC report, “The Value and Impact of Data Sharing and Curation”

Anne Wilson
April 21, 2016
Guiding Issues, Questions

- How is value to stakeholders defined/discussed?
- What are definitions/explanations/categories of repository stakeholders?
- How were similarities among data repos defined/discussed?
- How were differences defined/discussed?
- What, if any, were the metrics used to measure the values returned to stakeholders?
- What do they say about the reasons for caring about this topic?
What the report is

• A synthesis, summary and reflection, of combined findings of 3 prior investigations, by same authors, of value and impact of 3 data centers

• ESDS - Economic and Social Data Services

• ADS - Archeology Data Service

• BADC - British Atmospheric Data Centre
Reflections on:

- The methods that can be used to collect data for such studies
- The analytical methods that can be used to explore value, impacts, and benefits
- The measurable value, impacts, and benefits of the research data centers and the research data curation that they support
Similarities and Differences across the centers

- See Table 1, p. 7
- Differences in history, size, users
Methods

• Qualitative/economic analysis

• Quantitative analysis, non economic benefits

• Focus on users, depositors. Wider impacts are reflected, not measured

• Data collection included: prior reports, user surveys, registration and log analysis...
Quantitative/Economic Value

- See Figure 1, p. 9
- Left to right - transitions from measurable to estimated value
- Grey range - increasing impact
- Must estimate broader impacts
Types of Quantitative/Economic Values

- Investment and Use, a direct cost
- Contingent Value, estimated
  - If users can’t pay, a good measure is what they would accept to forego the service
- Efficiency Impact, estimated
  - e.g., impact on teaching, research
- ROI of hosting the data, estimated
  - e.g., data and services hosted and delivered over time
Quantitative Data and Approaches

• Table 2, p. 12, describes for each center how parameters were defined, data were collected and bases were established

• Initial decisions for each center:
  • who are users? depositors?
  • what is the basis for cost estimation?
  • salary estimates, etc.
Quantitative Results

- Figure 3, p. 14, shows incomparable comparison
- Over and over, “the results are not comparable”
- BUT, all results are favorable - there is clearly value!

- Customization of analysis is necessary
Quantitative Variables,
Appendix 1, p. 25

- Creation costs
- depositor costs
- operational budget
- investment value
- use value
- willingness to accept

- willingness to pay
- consumer surplus
- net economic value
- efficiency
- ROI on data creation

These can be reused!
Qualitative Value

• Data collected via interviews, w/ questions like “How severe would the impact if you could not access the data and services?”

• case studies - impact in terms of debate and media coverage of major social issues?

• only 3 performed for 1 study, no others

• The Keeping Research Data Safe (KRDS) Framework - “specifically developed for appraising qualitatively the benefits of digital preservation and curation of research data.”
KRDS

• See Figure 2

• A way to think about benefits, summarize findings

• Necessary inclusions:
  • discipline and community specific additions and variations
  • different relative prioritizations and ordering of benefits for each center

• Harder than qualitative! More manual, thoughtful, customized.
Qualitative Results

• Majority reported a major or severe impact if data and services were lost

• Allowed a stakeholder benefit analysis

• ADS stakeholder focus group revealed that impact study increased their recognition of the value of the center

• Benefit was known, but understanding of scope and value was increased
Important Overall Results

• Each approach independently gives only small view
• Each shows value, are complementary and mutually reinforcing
• While transferable, require significant customization to fit disciplinary and service differences
• The above limits cross-study comparisons
• Taken together they reinforce each other - the centers have a substantial and measurable positive ROI, and by facilitating additional use, that ROI can be increased.
Other Findings, Recommendations

- Assess at levels lower than ‘data center’
- Further develop these methods
- **Promote standardization of usage statistics**
- Track changes over time
- Study wider impacts - “it is in the uses to which the data are put after research use that substantial additional benefits can arise.”
Important Points

• The analysis was at the level of the data center itself. It could be done at a lower collection level. The level choice has implications

• Degree of center maturity matters

• Results are not directly comparable

• categories/classes/factors of/in analysis are identified, but significant customization is required
Guiding Issues, Questions Revisited

- How is value to stakeholders defined/discussed?
- What are definitions/explanations/categories of repository stakeholders?
- How were similarities among data repos defined/discussed?
- How were differences defined/discussed?
- What, if any, were the metrics used to measure the values returned to stakeholders?
- What do they say about the reasons for caring about this topic?
How is value to stakeholders defined/discussed?

- Both quantitatively, qualitatively
- It is necessary to factor in center specific characteristics in defining analysis details and implementation
What are definitions/explanations/categories of repository stakeholders?

- 3 centers covered
- Limited to users, depositors
- Report provides comments, reflections, estimates of wider impact, e.g. for funders
- Recommends even wider assessment of societal impact
How were similarities and differences among data repos defined/discussed?

- Carefully
- Repeatedly mentioned that results were not comparable
What, if any, were the metrics used to measure the values returned to stakeholders?

- See slide 11, Quantitative variables
What do they say about the reasons for caring about this topic?

- There is significant ROI!