



# Data Sharing and the E- infrastructure

*Michael Jubb*  
*Research Information Network*

STM Innovations Seminar  
4 December 2009

- 
- A vertical bar on the left side of the slide, transitioning from dark blue at the top to orange at the bottom.
- researchers' perspectives
  - skills and training
  - funding and institutional perspectives

# 1. Researchers' perspectives

## Constraints and disincentives

- large-scale collaborative projects and teams not the norm
- confusions over terminology
  - data, datasets, databases, digital objects, information
- creation, collection/gathering of data *not* usually the primary objective of research
- career rewards rarely come from sharing data
- resistance to *open* sharing of ‘intellectual capital’

## Ownership, protection and trust

- responsibility, protectiveness and desire for control over data
  - concerns about inappropriate use
- preference for co-operative arrangements and direct contact with potential users
- decisions on when and how to share
  - commercial, ethical, legal issues
- trust in other researchers' data?
  - *"I don't know if they have done it to the same standards I would have done it"*
- standardisation issues
- intricacies of experimental design and processes

## Curation and sharing

- data management and curation not yet adopted as standard practice
  - except in areas such as astronomy, bioinformatics, genomics etc
- other kinds of information more readily shared
  - software, code, tools, protocols etc
- challenges for service providers in meeting diverse needs of wide range of research groups
  - disciplinary and subject differences
  - subject knowledge
  - local, national and international
- relationships and engagement between researchers and information specialists

## 2. Training and Skills

*“This has been identified in every study as a major problem, both training researchers to be e-researchers, and training the people running the systems to deal with researchers, and to understand the technology”*

## Researchers

- *“...a lot of scientists don’t get information and structures at all. It’s not what they’re trained to think about.”*
- *“.....the idea of quality, provenance, and metadata about data is woefully inadequate in most science training.”*
- *“.....it’s not just about creating the infrastructure, it’s actually about creating the demand for the infrastructure, because.....there’s still a lack of demand.”*
- engagement between researchers and data management professionals
  - *“.....now we manage our data, whereas before we didn’t”*
- scalability in training and support, and in promoting cultural change

## Curators

- widespread concern about low numbers of people with specialist expertise
  - people from two kinds of backgrounds
    - library/information professionals
    - researchers
  - need for co-ordination of effort and funding for capacity-building
    - much depends at present on short-term project funding
  - lack of career structure
- *“Who’s training these people? We need training at the professional level for people who are actually going to run these data centres”*
  - *“...the career structure for those people with expertise is miserable, because the number of places they can work is not large, and the universities don’t treat them as key staff”*
  - *“So there’s a real danger of losing people to the private sector”*

### 3. Funding and institutional perspectives

## Leadership and co-ordination

- ▣ patchwork of well-established data centres
  - ▣ problems of co-ordination between different funding bodies
    - ▣ Research Councils, Higher Education funding bodies, JISC, universities
  - ▣ clarity about roles and responsibilities
    - ▣ piecemeal initiatives with limited take-up and impact
  - ▣ difficulty in the UK of establishing a single locus of responsibility
  - ▣ disciplinary and institutional dimensions of scale and complexity
- ▣ *“...we need a more co-ordinated strategy and real leadership to take things forward.”*
  - ▣ *“.....it’s very easy in the current framework to pass the buck and do nothing.”*
  - ▣ *“It would be good to have some national view of process for deciding who is responsible. It’s not clear what routes you have to go through to decide should we do this.”*
  - ▣ *“Things are funded in silos. So I just don’t think there is really a national strategy.”*

# Co-ordination and linkages beyond the UK HE/research sectors

- **Government and public sector information**
  - need to work in parallel on issues such as standards and interoperability
  - co-ordination but not monolithic approaches
- **commercial sector**
  - some scepticism, and relatively few examples or models of public-private partnerships
- **international infrastructure**
  - need for greater clarity within UK in order to optimise its role in international initiatives
- *“.....if you pick off these issues individually then you get one answer, but you may also want to look at how they operate as a whole system.”*
- *“.....there’s a real need to work out a research compact between the private sector index and data holders, search engine providers, and the public scientific sphere. That’s an urgent piece of work”*
- *“So you do need a strong champion, and it’s got to be someone who’s capable of operating at a very senior level.”*
- *“.....working with international partners helps to set the aspiration nationally.....and it begins to give funders a mandate to ensure that some of the funding is there”*

## Drivers and Leaders

- agreement that development of the data infrastructure should be “driven by the science”
- securing engagement, commitment, time and resources from research leaders
- dangers of “solutions looking for problems”
  - *“.....as soon as you create a facility that isn’t focused on the science, it develops a life of its own.”*
  - *“.....you have to work pretty hard to demonstrate there’s a business case for reuse of data.....there’s no point in paying to curate and store data if nobody ever does use it again.”*
- need for careful management of relationships between specialists and researchers

# Top-down and bottom-up

- bottom-up
  - develop services locally in response to what researchers themselves want
  - develop tools and environments within universities to equip the research community with appropriate processes and skills
- top-down
  - establish national body/programme to catalyse change required for sustainable and ubiquitous service
- UK position
  - currently lacks structure for developing effective top-down approaches with appropriate engagement from leaders across wide range of disciplines
  - preferred solution seems to be small-scale initiatives working together
  - will that work and will it lead to sustainable services?



Thank you

Questions?

*Michael Jubb*

*[www.rin.ac.uk](http://www.rin.ac.uk)*