

# Open Data for Open Science

Geoffrey Boulton

ALLEA, Rome 2012

Some emerging conclusions from a  
Royal Society Policy Report:  
*“Science as an Open Enterprise”*



THE ROYAL SOCIETY

# Why does open data matter?

- Maintaining scientific self-correction (closing the concept-data gap)
- Exploiting data-intensive science
- The potential of linked data
- “Data is the new raw material for business”
- Exposing malpractice and fraud
- Stimulating citizen science
- Engaging with citizens in “public interest science”
- Aspiration: all scientific literature online, all data online, and for them to interoperate

# Openness of data *per se* has no value. Open science is more than disclosure

It must be communicated effectively. It must be:

- Accessible
- Intelligible
- Assessable
- Re-usable

]**METADATA**

Only when these four criteria are fulfilled are data properly open

Metadata must be audience-sensitive

Scientific data rarely fits neatly into an EXCEL spreadsheet!

# Boundaries of openness?

- Legitimate commercial interests
- Privacy (note that anonymisation is impossible)
- Safety & Security

But the boundaries are fuzzy & complex

# Maximising the benefits of open data

- **R & D on software tools** (semantic linking, citation, persistent identifiers, sorting etc: note that the ICT industry is often way ahead)
- **Institutional responsibility for the knowledge they create** (cumulative small science data > cumulative big science data)
- **Data scientists** (they are being trained, and the commercial demand is large)

“Big Iron” is a national infrastructure priority

“Big data” is a science priority – **the big costs are people and software, not computers**

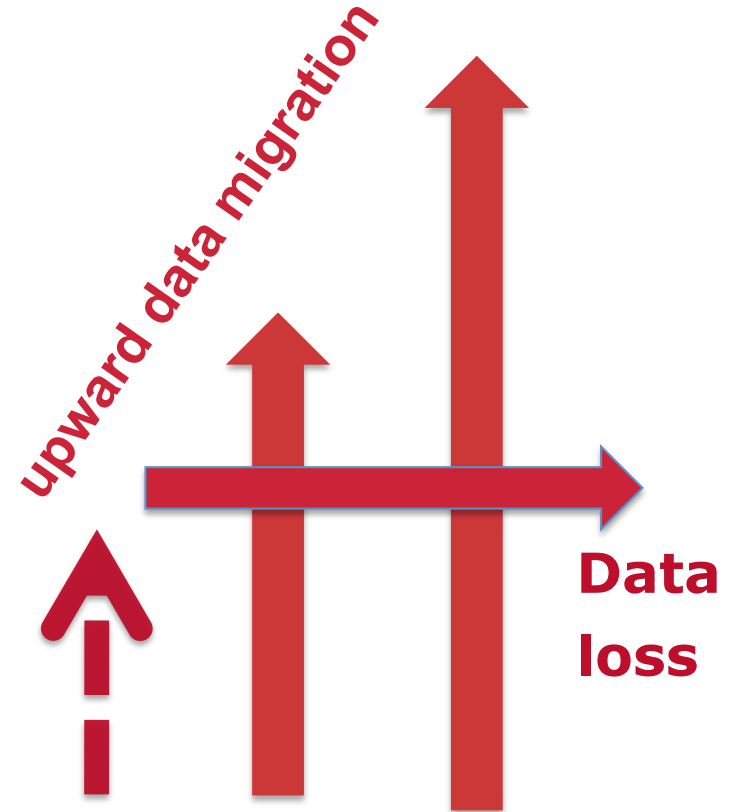
# Levels of data curation

**Tier 1** – International databases

**Tier 2** – National  
(e.g. Research Councils)

**Tier 3** – Institutions  
(Universities & Institutes)

**Tier 4** – “Small science” researchers  
& research groups



**Financial sustainability?**

# Actors in stimulating change

- Scientists: persuading them to act
- Employers (universities/institutes)
- Funders of research (the cost of curation is a cost of research)
- Publishers of research
- Business – exploiting the opportunity
- Government – is it important? If so, do we need to act?

# Responsibilities

**The role of academies  
& learned societies?**