Improving Data Management Practices of Researchers by Using a Behavioural Framework

Malcolm Wolski and Joanna Richardson
The “three V’s”, i.e. Volume, Variety and Velocity of the data coming in is what creates the challenge.

**Volume**
- >3,500 North America
- >2,000 Europe
- >250 China
- >400 Japan
- >500 Middle East
- >200 India

Amount of Big Data stored across the world (in petabytes)

**Variety**
- People to People: Netizens, virtual communities, social networks, web logs...
- People to Machine: Archives, medical devices, digital TV, e-commerce, smart cards, bank cards, computers, mobiles...
- Machine to Machine: Sensors, GPS devices, bar code scanners, surveillance cameras, scientific research...

**Velocity**
- 2.9 million emails sent every second
- 20 hours of video uploaded every minute
- 50 million tweets per day

8,000 billion is the potential annual value to Healthcare

**Value**

<table>
<thead>
<tr>
<th>Industry</th>
<th>Productivity Increase</th>
<th>Sales Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retail</td>
<td>49%</td>
<td>$9.68</td>
</tr>
<tr>
<td>Consulting</td>
<td>39%</td>
<td>$5.08</td>
</tr>
<tr>
<td>Air Transportation</td>
<td>21%</td>
<td>$4.3B</td>
</tr>
<tr>
<td>Construction</td>
<td>20%</td>
<td>$4.2B</td>
</tr>
<tr>
<td>Food Products</td>
<td>20%</td>
<td>$3.4B</td>
</tr>
<tr>
<td>Steel</td>
<td>20%</td>
<td>$3.4B</td>
</tr>
<tr>
<td>Automobile</td>
<td>19%</td>
<td>$2B</td>
</tr>
<tr>
<td>Industrial Instruments</td>
<td>18%</td>
<td>$1.28</td>
</tr>
<tr>
<td>Publishing</td>
<td>18%</td>
<td>$0.8B</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>17%</td>
<td>$0.4B</td>
</tr>
</tbody>
</table>

40% projected growth in global data created per year

$55 5% projected growth in global IT spending per year

The estimated size of the digital universe in 2015 was 1.8 zettabytes. It is predicted that between 2009 and 2020, this will grow 44 fold to 35 zettabytes per year. A well defined data management strategy is essential to successfully utilize Big Data.

Let’s not forget the long tail of research data

- Small independent research efforts fall in the long-tail of the distribution
  - Most of this (such as siloed databases, null findings) is unpublished
  - These dark data hold a potential wealth of knowledge

http://www.slideshare.net/SusannaSansone/sansone-statement-bdebatetue11nov2014
“DATA IS THE NEW GOLD”

http://www.slideshare.net/itweekend/building-a-data-driven-organization
Open Access Mandates

(remind me, what’s the objective again?)

http://www.slideshare.net/tobygreen/freemium-open-access-publishing-learning-to-let-go
Sharing Clinical Research Data: An IOM Workshop

October 4-5, 2012
National Academy of Sciences
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

http://www.slideshare.net/marilynmann/sharing-clinical-research-data-an-iom-workshop
Collaboration and co-ordination: Ecosystem science research cycle(s)

- **Ecosystem Science**
  - **Enhanced ability to revise, question and expand knowledge**

  - **Knowledge gap: research questions**
  - **Proposal and planning**
  - **Data collection, verification, quality assurance and control**

- **Increased effectiveness**
  - **Storage, preservation and discoverability of data**
  - **Data analysis, integration and synthesis**
  - **Research output: new data and publications**

- **Efficiency gain**
  - **Enables large scale and coordinated data collection, sharing and multiple re-uses**
  - **Data + meta-data, licensing**

http://www.slideshare.net/TERNCOMMS/terns-data-haring-presentation-at-try-workshop
Risk Management

http://eresearch.uws.edu.au/blog/author/alf/
Managing Research Data: New Roles for Librarians

Constance Wiebrands & Julia Gross

http://www.slideshare.net/jmcgross/managing-research-data-new-roles-for-librarians
IT TAKES TWO TO TANGO - BY SNOWPEA

IT TAKES TWO TO TANGO

TANGO!!!

http://www.toondoo.com/cartoon/4856032
Why behavioural models

RESEARCHERS

- 60.7% reported a formal data management plan but 39% reported backing up their data monthly or less
- Few researchers, especially early career, think about long-term preservation of their data
- Demands of publication output overwhelm long-term considerations of data curation
- Metadata and documentation are of interest only if they help a researcher complete his or her work
- Many researchers expressed concerns surrounding the ethical reuse of research data
- Lack of time to conduct basic organizational tasks, let alone time to research best practices or participate in training sessions
- Many sceptical of long term interest in their data

What are behavioural models

- Methods developed for studying behaviour
- Some are targeted to population groups or problems
- Two categories of interest to us
  - Major Theories of Individual Behaviour / Change
  - Major Social and Technological Theories of Behaviour / Change

http://blogs.scientificamerican.com/mind-guest-blog/2013/06/14/better-behaved-behavioral-models/
## Example of Theories of Individual Behaviour / Change

<table>
<thead>
<tr>
<th>Theory</th>
<th>Major Tenets</th>
<th>Comment</th>
</tr>
</thead>
</table>
| Theory of Planned Behaviour (TPB)               | • Examines the link between intention to act and performing a behaviour.  
• Intention is determined by an individual’s attitude (belief and values about the outcome) and subjective norms.  
• Behaviour is also determined by an individual’s perceived behavioural control. | Useful for reinforcing the need to present information in a way which helps shape positive attitudes toward behaviours. |
Social Cognitive Theory (SCT)

- Examines how behaviour, personal and environmental factors interact to determine human functioning.
- Major elements which may intervene include self-efficacy, outcome expectations, reinforcements (something that increases / decreases likelihood a behaviour will continue), and observational learning (acquiring behaviours by observing others’ behaviour).

Useful for looking at resources which could raise self-efficacy, determining whether incentives are required, and recognising environmental constraints that might deter behaviour change.
Efficacy

Attitudes  Benefits  Reactance

Intentions  Subjective

Cues  Barriers  Response

Action  to  Self  Norms  Fear  Threat
A-COM-B framework for understanding behaviour
adapted from COM-B model

Attitude

BELIEFS

AFFECT (FEELING)

BEHAVIORAL INTENTIONS

http://www.consumerpsychologist.com/cb_Attitudes.html
Capability is the psychological or physical ability to enact the behaviour (Michie et al, 2011, p. 4). This **perceived** capacity to adopt a behaviour (their self-efficacy) is fundamental to a person taking any action to change their behaviour.
Opportunity is defined as “all the factors that lie outside the individual that make the behaviour possible or prompt it” (Michie et al., 2011, p. 4)

http://evolution.berkeley.edu/evolibrary/article//success_09

Motivation is defined as “all those brain processes that energize and direct behaviour, not just goals and conscious decision-making. It includes habitual processes, emotional responding, as well as analytical decision-making” (Michie et al., 2011, p. 4).
1. Changing behaviours or attitudes?
2. Attitudes and responses change over time
3. Start with the individual – not the service/product
4. Start with the cohort first
5. Where is the opportunity/motivation?
6. Perception of capability is important
The Importance of a CRM to target cohort-individual

Trial with a small cohort eg a research centre

Target HDRs

“Trialling this framework provides an opportunity for our Discipline Librarians to systematically approach and advise researchers of our institutional data storage options and the benefits of tailored storage solutions “
Questions