TURNING BIG SHIPS (DATA) ON A DIME

CHANGE MANAGEMENT AND DATA SENTENCING

Dave Connell, Sandra Ennor, Rhys Francis,
Nicholas McPhee, Jaye Weatherburn, Lyle Winton

Agenda

Introduction

- Dave Connell – Australian Antarctic Division – Commonwealth of Australia (10 mins)
  - **How to easily capture metadata – a government perspective**
- Sandra Ennor – Monash University (10 mins)
  - **Research Data Culture**
- Lyle Winton & Jaye Weatherburn - University of Melbourne (10 mins)
  - **Digital Preservation Strategy**

Discussion (30 mins)

Introduction

Australia's initial national eResearch strategy was developed in 2006. The feedback received was very simple

yes we need advanced computing…. yes we need advanced networks... but

it's all about data

ANDS and NeCTAR and RDS were born, and today we seek to build a data commons.

The impact of data being digital as digital technologies evolve is massive.

- It allows 'observation' to be made efficient, repeatable, scaled-up, comprehensive, recorded and shareable
- It allows us to see things unfold over time, making research into 'processes' and 'change' possible
- It allows us to see populations and their properties and thus discern patterns otherwise invisible
- It better connects observation, theory, modelling and experiment and connects research and the real world
- It enables transformations from connectivity, interactivity, automation and machine intelligence in research

Introduction

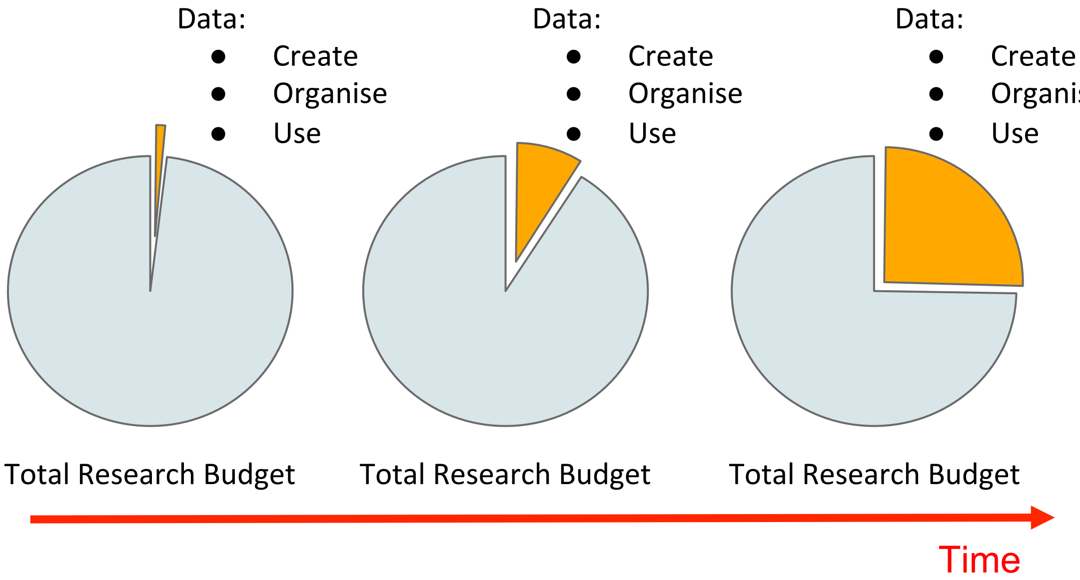
The research world of 2018 is much more data savvy and also becoming more research software savvy.

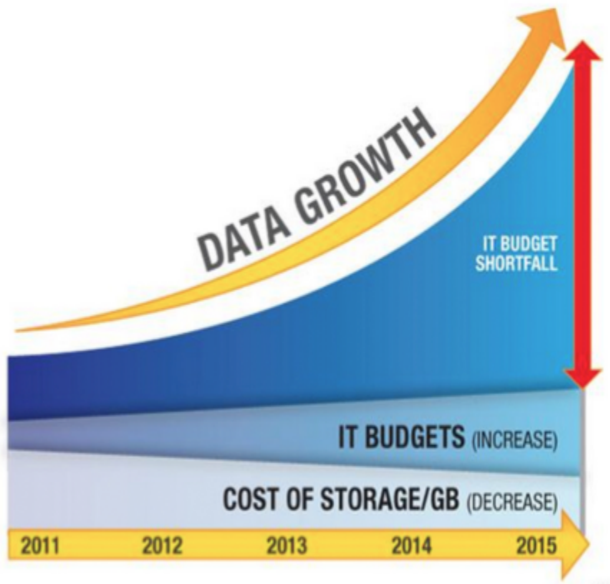
Today there are four key data infrastructure requirements [from researchers]:

- We need better data
    - More advanced instruments and sensors, more and better metadata, access to non research data
- We need to keep and share data better
    - Higher quality collections, FAIR for open data, and easier collaboration around controlled access data
- We need to use data better
    - Access to world-leading best-practice digital data methods and related data (a global methods commons)
- We need to relate data better
    - Better integrations of larger holdings of more complex data, for new science and new research

And then there are institutional, legislated and social goals and requirements to consider

Introduction - Is it all good news then?





It is in fact incredibly easy to create mountains of unimportant digital data.
Knowing what data is valuable compared to those mountains of unimportant data is hard.
And, once a significant share of budget is applied, data will have to live within its means.
When we reach that point, we will need to turn from “keep & curate it all” to something else - fast.

Discussion

The draft statement of intent reads as follows - we need solutions that:

- Deliver cost effective management and curation processes
  - that support societal, institutional and researcher data goals (concurrently)
- Use technology and deploy infrastructure to minimise additional work
  - for researchers (but everybody really), reducing the ‘workload’ caused by data would be even better
- Focus curatorial investment on the most valuable data
  - prioritised by institutions to enhance societal contribution and reputation
  - prioritised by researchers to enhance research impact
- Hold a larger share of the overall data in less expensive services
- Contain the total cost
  - by containing the total amount of data stored and curated

Discussion - will we tackle this together?

- What would an *affordable research data culture* look like
  - where the behaviour of researchers and research services optimises institutional and research outcomes, and balances cost and impact?

- When do we know data is destined to be made openly accessible or contributed to referential data sets, collaborated on, protected, commercialised, retained in defence of research claims, or discarded?

• How do we sentence data into different outcomes (open access, protected, defensive, discardable) and into different performance tiers of storage and access, with certainty over time? • Can we merge the tiers of storage (HPC, cloud, desktop and archive) seamlessly in a way where research practice optimises (for us) across those tiers? • Can we create data that is 'inherently FAIR' - where no additional human labour is required?	• How does stewardship need to change as data undergoes a transition into longer term retention cycles? • What resources can be put in place to build and support an improved Research Data Culture? • How does Research Data Management relate to Research Data goals? • How do we implement policy/governance?