

What large clinical data sets exist across the member organisations of MACH, Melbourne and how can we collate and curate them to maximise research outcomes?

Associate Professor Dougie Boyle
Academic Convenor, Petascale Campus Initiative
Faculty of Medicine, Dentistry and Health Sciences
The University of Melbourne

The Project

- Datasets within hospitals and research institutes are often not well known or characterised. Hidden gems may be out there
- These may be clinical databases or research-specific resources
- As part of considering data science uplift, we wanted to understand how we could better identify such resources on an on-going basis to maximise research potential
- FAIR Focus: Finding datasets and promoting reusability
- We aimed to utilise simple REDCap surveys as the principle knowledge discovery tool with follow-up interviews with identified key individuals
 - A simple survey to begin with: What is your dataset and who's the contact person
 - A more in-depth survey to capture more complete dataset metadata

Setting & Wider Goals

Melbourne Academic Centre for Health: The MACH is a National Health and Medical Research Council (NHMRC) recognised Advanced Health Research and Translation Centre (AHRTC)



Wider goals:

- Provide lessons learnt to our Australian research data community
- Evolve models and tools that allow for on-going dataset discovery

Results

- It took nearly three months before data started to flow
- The collection now stands at 180 databases discovered with this increasing fast
- The data is a diverse mix of RCT's, interventions, observational datasets and clinical datasets
- Almost none were known to the PI in advance
- Although hard to initiate and slow to start, the initiative success means it is being funded 0.4EFT for the next year via the PCI and the Faculty and looks to be a foundational component for our research strategic planning going forward
- We have a long way to go to turn this into an on-going, live collection
- The dataset descriptions and metadata will be available once we have had time to curate it and obtain permissions for open sharing after collaborating institution review

The Key Issues

- **Engagement**

- The institutions have very formal processes regarding engagement for quality improvement processes and research
- Regardless of the fact that a simple survey did not fall under the remit of quality improvement or research, ethics approval was still demanded by many
- Ethics approval from The University of Melbourne was accepted by many organisations but not all
- One organisation requested the PI complete Good Clinical Practice training

- **Time**

- Many processes were out-with our control in-terms of speed of approvals processes and contacting key individuals at each institution
- Including plain English language statements, the number of documents needing submitted to request permission for such a survey could be up-to nine
- In some cases, weeks of email and phone communication was required

- **Cost**

- Most organisations did not charge but one hospital charged \$1,100 as an administration fee for research committee and ethics review

Lessons learnt

- Engagement takes time – we required a 3-month initial time window to engage with our collaborating organisations utilising a highly experienced BA resource
- Continued longitudinal follow-up and engagement will be required to ensure we have engaged fully within the organisations
- Commence an ethics application through your University immediately at the outset even if advice is that is not required
- We are still learning how to turn this into datasets that can contribute to Research Data Australia – we have started this journey so please engage with us if you wish to undertake similar dataset discovery

Acknowledgements

- ARDC for project seed funding
 - Thanks to Richard Ferrers in particular for on-going advice and support
- University of Melbourne Petascale Campus Initiative
 - Additional project funding for the initial project duration
 - 0.2EFT support for the next year
- The Centre for Clinical and Population Health Informatics (Wendy Chapman)
 - 0.2EFT support for the next year
- Melbourne Academic Centre for Health (MACH)
- Thanks to all the contributing MACH collaborating organisations
 - Research offices
 - Nominated organisational leads
- Ursula Soulsby – the BA who has worked tirelessly and with endless patience to make this a success

