SENTENCING FOR MURDA

e-Research Australasia Conference

Brisbane, 2017

Catherine Nicholls (Records Manager, eSolutions)

Main co-contributor - Nicholas McPhee (Research Data Management Specialist, MeRC)

Other credits: Steve Quenette (Deputy Director, MeRC), David Groenewegen (Director, Research, Library), Stephen Dart (Research Storage Manager, MeRC) and Neil Dickson (Research Infrastructure Manager, Library).

MONASH University

Overview of the issue

A lot of data is hanging around, some of it is obsolete, redundant, due for deletion…

Metadata often missing… Hard to ‘police’ (manage) it over time.

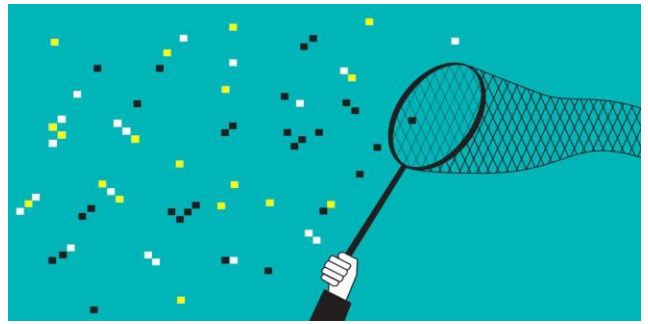
Data gets up to ‘no good’…

Costs a lot to maintain & can make high value data harder to locate…

MONASH University

Sentencing for MURDA – a way to tackle the data blob problem…

- Background
- Sentencing
- MURDA (Monash University Research Data Archive)
- What have we learnt?
- Next steps



Data… We all have a lot of it.

MONASH University

Monash University at a Glance

Est. 1958

Australia's *largest* University

63,000+
Students Worldwide

16,000+
Staff Worldwide

7,000+
Research Staff Worldwide

Global Reach

Italy *China* *India* *Malaysia* *South Africa* *Australia*

$$

$400m+
Research Income 2014

MONASH University

Background – MAWG (and me)

MAWG – Monash Agent Working Group

Members from ***many functions*** vs advisory to any one function.

Practicing members from the following institutional functions: Information Management (eSolutions) [me!], Infrastructure Services (eSolutions), Library, Research storage, cloud and HPC (e-Research), and Service Office (eSolutions).

Leaders from Library, eResearch and eSolutions.

MONASH University

NCRIS change agenda translated into the institution

Original mandate:

The institutional response to the NCRIS 2007 (NeCTAR/RDS[i]) change mandate of accessibility and collaboration. The goal was to accelerate Monash researchers' adoption of this change and resource as it relates to data.

Current mandate:

Accelerating Monash researchers' adoption of accessibility, collaboration and data management (e.g. now getting into things like data governance).

MONASH University

Definition of sentencing

Monash University has its own Retention and Disposal Authority (RDA). Based on Public Record Office of Victoria (PROV) sentences.

Sentencing is the process of matching information held by the organisation to a specific class of a records authority.

This helps determine the value of the information and how it should be managed throughout its lifecycle.

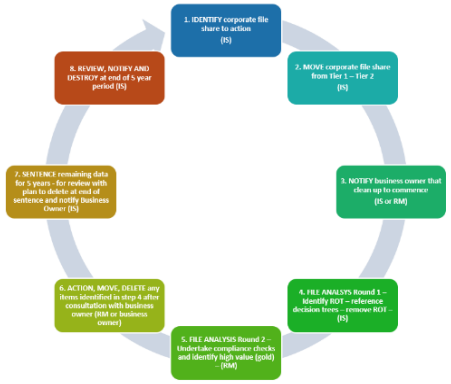
MONASH University

MURDA and data lifecycle strategy

Before I talk about MURDA specifically, need to mention a related project for context…

Data Lifecycle Strategy (corporate and research data)

* Decision trees
* Some file analysis tool assessments
* Also led to some specific sentencing work – e.g. some 45TB redundant ‘personal drive’ data, data in MURDA, etc...



MONASH University

MURDA – the basics

MURDA is a model that enables sentencing to occur.

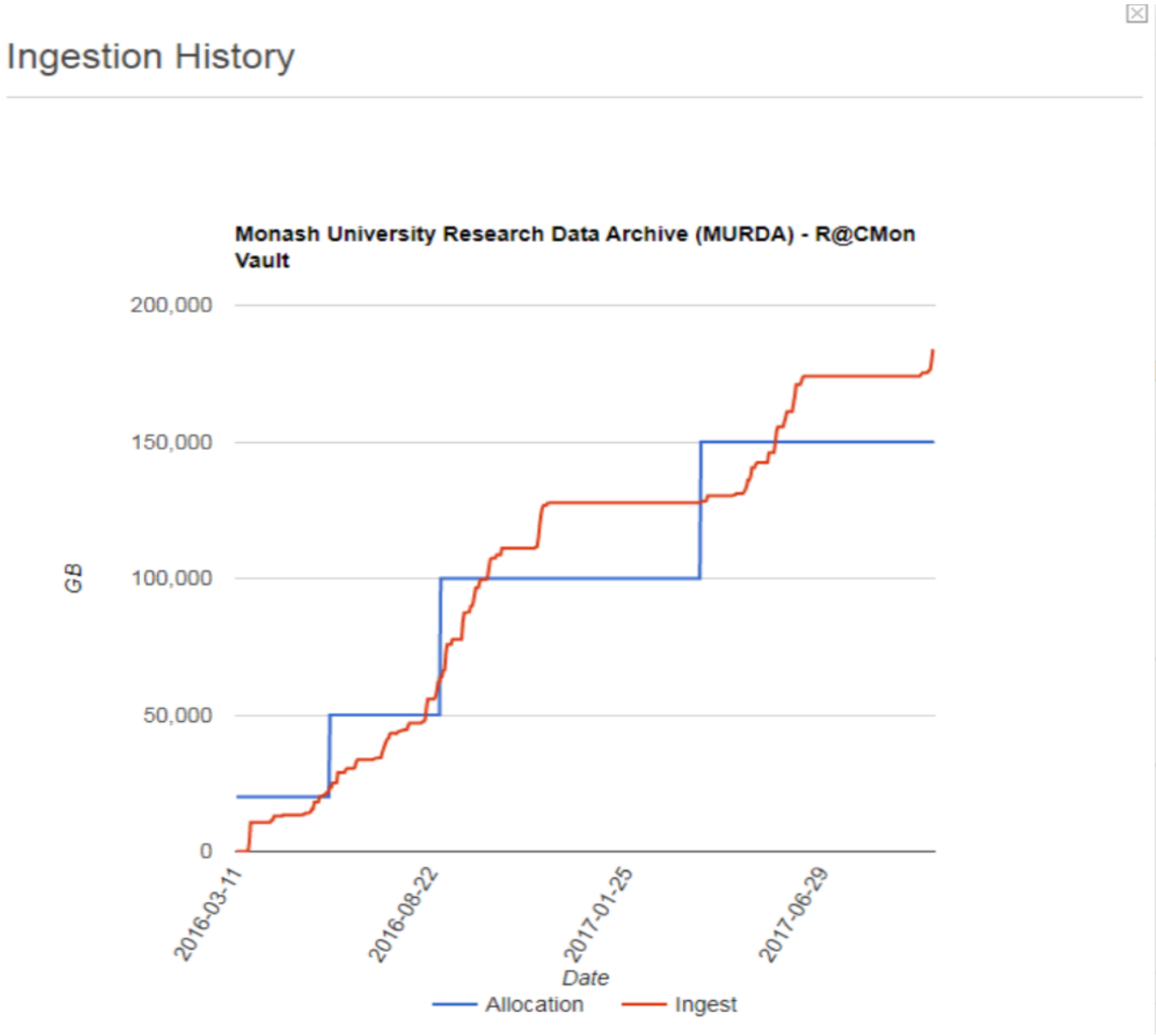
MURDA combines storage with a front-end database that captures metadata about mostly ‘retired’ data.

For the those who are interested MURDA has:

- 233 retired projects and is still growing;
- 184TB of data that consumes ~100 media; cartridges (2 copies, backup and HSM);
- Only about 9TB is online (~5%);

MONASH University

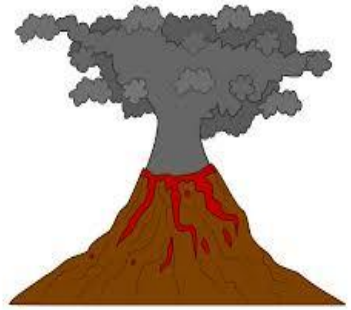
Some MURDA stats…



MONASH University

How does research data get into MURDA?

- Data may need to be moved (sometimes urgently) from an existing location. Stuff happens.
- Researcher/research group alerts e-Research that project closing down – can’t destroy data yet.
- e-Research staff become aware data has ‘no place to go’ through their day to day dealings with researchers.



MONASH University

Before data is MURDA'D

The data is INTERROGATED (intensely)!

Who is the owner?
Is it orphaned data?
For how long does it need to be retained?
When was it last accessed?
All documented within MURDA.

MONASH University

Sentencing for MURDA.

PROS 16/07 - 3.3.2	Data and datasets created as part of research activities within the institution, which are of regulatory or community significance.	PERMANENT	**Permanent Archive**
PROS 16/07 - 3.3.3	Data and datasets created from clinical trials as part of research activities within the institution. Excludes data and datasets included in class 3.3.2	TEMPORARY	**Destroy 15 years after completion of research activity**
PROS 16/07 - 3.3.4	Data and datasets created as part of research activities within the institution which involve minors. Excludes data and datasets included in class 3.3.2	TEMPORARY	**Destroy 15 years after child reaches the age of 18**
PROS 16/07 - 3.3.5	Data and datasets created as part of research activities within the institution. Does NOT include data created for specific research activities for which additional regulatory requirements apply, including: clinical trials, gene therapy and research involving children. Excludes data and datasets included in class 3.3.2	TEMPORARY	**Destroy 5 years after completion of research activity**

SOURCE: PUBLIC RECORD OFFICE VICTORIA

MONASH University

Process for sentencing in MURDA

MeRC staff and University Records Management team assess available metadata and sentence data.

As per existing processes, University Archives Manager will then:

- Compile lists of data with attached sentences
- List will be sent to senior Faculty representative (or equivalent) with a set timeline for action attached.
- Data that is eligible for legal destruction will be destroyed.

MONASH University

Sentencing in MURDA

Known exceptions:

- DOIs?
- Some clinical datasets?
- Instrument-based research data?

MONASH University

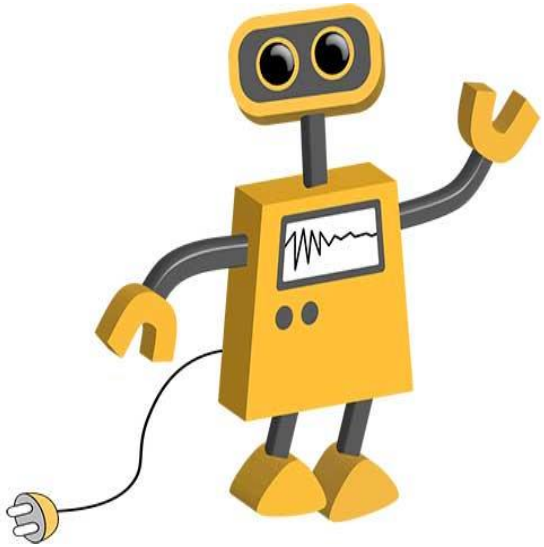
What have we learnt?

- The old chestnut - we want it kept – just because…
- Providing a ‘service’ & ‘infinite space’ versus ‘capacity/maintenance/cost issues’ & ‘focusing high end resources on high value data’.
- Orphaned data.
- “Interrogation” process resource-intensive.
- When no one wants to make a decision?
- Destroying data is scary! Easier to just ‘archive it’…

MONASH University

What else have we learnt?

- MURDA as a model – allows for the ‘unplug and see’ approach.
- The researcher is always going to be in the best position to sentence data.
- Important to ascertain who business owner of data is, whether they are currently ‘active’ and ‘last accessed date’
- Metadata needs to be retained with the data – or linked to it.
- SENTENCE AT CREATION



MONASH University

Next steps…

End goal ➡ engage tools that allows researcher to appraise data themselves (easily).

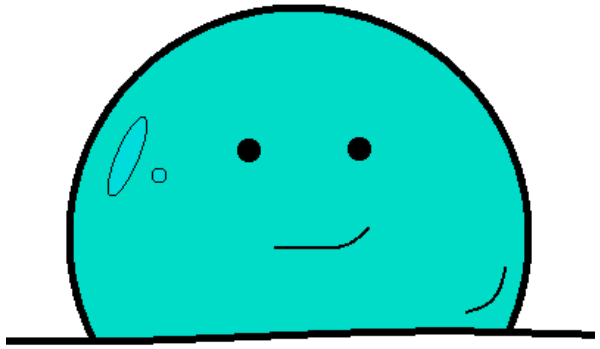
Harness existing metadata better (e.g. Human Ethics records the retention period).

Education. Disposal is *not* a swear word. When done legally and compliantly actually saves $$ for the right things, removes burden of maintaining junk/redundant (blob) data.

MONASH University

Full Credits

Catherine Nicholls - Catherine.Nicholls@monash.edu
Nicholas McPhee - Nicholas.McPhee@monash.edu
Content contributors:
Steve Quenette - Steve.Quenette@monash.edu
Stephen Dart – Stephen.Dart@monash.edu
David Groenewegen - David.Groenewegen@monash.edu
Neil Dickson - Neil.Dickson@monash.edu
Adrian Tritschler - Adrian.Tritschler@monash.edu
David Lam – David.Lam@monash.edu



*Please note I will be available at the **FAVeR Federation for the Advancement of Victorian eResearch** booth directly after this presentation.*

MONASH University

Image Credits

- Slide 1-https://d2v9y0dukr6mq2.cloudfront.net/video/thumbnail/TOEwt0C/court-law-justice-litigation-concept-with-gavel- and-hammer_ey7bkhxbe__S0000.jpg
- Slide 2- https://media3.giphy.com/media/IQYoP7rv3Qz6g/giphy_s.gif; https://pbs.twimg.com/profile_images/3788000000799872595/266b744ab3aa1c20b41416159e6506ff.jpeg; https://sf-ugc.wpcdn.pl/user_photos/537576a3c95381_91679293.jpg;, https://static1.squarespace.com/static/5225e6a0e4b02f25c1df67cf/t/58c026a73a041163d87bbf8f/1488987830133/
- Slide 3-https://pbs.twimg.com/media/DCigcFWXgAEWyHU.jpg
- Slide 4-Steve Quenette, Monash Unviersity.
- Slide 5- https://cinematice.files.wordpress.com/2012/10/inclusion-equipe.png?w=1200
- Slide 6 - https://beta.theglobeandmail.com/resizer/f7YKTH94oYZRQA6sYeZ7MdW-Y8Y=/3500x0/arc-anglerfish-tgam-prod-tgam.s3.amazonaws.com/public/3TKPDJVGW5CTNPAEY3FQ46KZ5I
- Slide 7- https://lh3.googleusercontent.com/-4VK15ycvhUI/Vszy2NtMf2I/AAAAAAAAM6k/r1JT3JzrahU/w530-h478-n/cartoon-judge-009.jpg
- Slide 8 - Catherine Nicholls, Monash University.
- Slide 9 -

https://1.bp.blogspot.com/-aqJ3nDjxeTs/Vzk9a251z9I/AAAAAAAAC5U/qkiDo72OdasJtgbOPVX9NXKpZxx5QGSmQCLcB/s1600/Blob.jpg

MONASH University

Image Credits

- Slide 10 - Stephen Dart, Monash University.
- Slide 11 - https://agendacopiilor.files.wordpress.com/2017/09/21370941_1857365761243694_4044025657784124642_n.jpg?w=350&h=200&crop=1
- Slide 12 - Catherine Nicholls, Monash University (mashed together image).
- Slide 14 - http://amothersshadow.com/wp-content/uploads/2015/04/Monster.jpg
- Slide 15 - https://i.pinimg.com/originals/68/aa/3f/68aa3fa95bfe726389df9f453eafb7b6.jpg
- Slide 16 - https://static1.squarespace.com/static/5225e6a0e4b02f25c1df67cf/t/58c026a73a041163d87bbf8f/1488987830133/, , https://images.emojiterra.com/google/android-oreo/512px/1f92f.png
- Slide 17 - https://fcit.usf.edu/matrix/resources/tech-cartoons/
- Slide 18 - https://static1.squarespace.com/static/5664c8e6e4b073dc48251fd7/5665111fe4b055b21a442993/56651120e4b02b0e954142cb/1449464097403/blob-walk.gif
- Slide 19 - https://www.weasyl.com/~blobskin/submissions/864316/my-happy-blob
- Crediting images found online correctly can be difficult. Best attempts have been made to note the source of the images.

MONASH University