The research publications workflow at Deakin University
Or, our transition from a manual to an automated workflow
Prepared for the CAUL Repositories Community Day
June 2017
About Deakin

- Young institute around 50 years old
- @1500 researchers working across 4 campuses, 4 faculties, 4 research institutes and 13 strategic research centres
- Expanding our research capabilities and reputation
- One of Australia’s fastest growing research universities - research funding increase from A$4.5 million in 1997 to A$47.2 million in 2015
- 2015 ERA rankings: 89% of Deakin research rated at or above world standard
- 2015 Times Higher Education 100 Under 50 ranked Deakin University 45th in the World
Deakin Research Online

- Created in 2008 / uses Fez/Fedora software
- Mixed repository of open access and dark archive material
- Broad collection policy based on the philosophy of “we’ll take anything created by a Deakin researcher”
- Publications source of truth > publication data fed from the repository to Research Master then onto Staff profiles
- Strong relationship between Faculty, the Library, and the Research Office
Our publications workflow from 2008-2014 was based on manual processes involving record by record review of publication metadata and outputs.

- **Faculty**
  - Views each and every publication before assigning a HERDC classification

- **Library**
  - Adds additional metadata for discovery, investigates OA and publishes the record

- **Metadata fed to Research Master**

- **Metadata fed through from RM to Staff profiles**

*This was manageable when we were dealing with @100 records input manually each week.*

Deakin Research Online (Fez/Fedora)
Introduction of Elements, late 2014

- The Library is the business owner of Elements
- Elements creates profiles for researchers, then harvests publication metadata from sources such as Scopus, WoS, etc.
- Metadata is fed through to DRO, then onto RM and Staff profiles
- We adapted the HERDC module to control workflows based on the old model of manual review
- Searches were turned on in late 2014, and Elements began harvesting data
- Our backlog increased dramatically ...
#Records harvested by Elements, vs. #Records published to the repository
Why a backlog?

• No guidelines
  • No clear guidelines to assist with prioritising material
  • No guidelines on who owns the workflow, and who is responsible for what, as well as timelines

• Administratively cumbersome
  • Everything is nominated for HERDC > everything goes through Faculty
  • Letters to the editor receive the same attention/checking as peer reviewed articles
  • Faculty view all pdfs to verify quality, regardless of publication source
  • Liaison with the researcher to obtain pdfs leads to hold ups

• Inefficient workflows and metadata feeds

• Work from a priority list, no guaranteed turnaround time
  • ERA / Staff promotions / New researchers / 2011-2017 publications
Impact on researchers

> Staff profiles out of date
> Delays in making Deakin research discoverable
> Impacts on citation rates
> Complicated process that researchers find confusing

I have accepted my publication – why is it not on my Staff profile!?!}

I am new to Deakin ... this publications workflow is so complicated!!
Solutions and improvements
Our values & assumptions

There is value in retaining Faculty expertise in the pubs workflow
Research publications need to be vetted and classified as they come thru Elements
Non reportable outputs don’t need the same high level of processing as reportable publications
There needs to be a “One Deakin” approach toward the processing of publications
The repository is important and is valued:
- to Preserve research
- to make it Discoverable
- to make research Openly Accessible (to increase citation rates)
Consultative workshops, 2016-17

• Combined Library, Faculty, and Research Office workshops
  • To provide a forum to voice opinions and concerns
  • To map out and identify variations between the Faculties in their approach

• Workshops including the Library, Deakin Research and Leonie Hayes from Symplectic
  • To review metadata-based workflow and systems triggers in the Elements > DRO > Research Master data feeds
  • To re-engineer workflows and data feeds
  • To review harvested metadata and determine ways we can automate our processes whilst ensuring quality
New “Smart ingest” approach to harvested metadata

• Journal articles should be accepted as C1 or C1.1, and meet the definition of research without additional checking if:
  • The publication was found by Scopus/Web of Science
  • The publication is of type “article” in SCOPUS/Web of Science
  • The publication is more than 2 pages long
  • The parent journal is in SSCI/SCIE or in SCImago Q1 (as per Where Should I Publish)
  • The title does not include the words “Review”, “Editorial”, “Case Study” or other known stop words likely to indicate the item does meet the definition of research.
• A spreadsheet of claimed journal articles in Elements is generated nightly and output to a shared Library/Faculty wiki site
• The report is generated by a custom PHP script running on DRO
• The script queries all the journal records in Elements, does an ISSN match against a pre-loaded dataset that includes various quality matrices (e.g: SCImago journal score, quartile, impact factor, number of pages, etc.)
• Faculty staff download the spreadsheet and apply filters to identify journal articles that automatically qualify as C1s
• The spreadsheet has been greeted with enthusiasm by most Faculty staff
• It means they don’t have to view every publication that’s coming through, and can confidently assign a C1 classification to articles that meet the criteria
• Processing time has been doubled using the new techniques
Next steps: working with Symplectic to further automate our workflow

• The spreadsheet is useful for automatically identifying C1s, but the Elements interface requires Faculty to process records individually

• Working on utilising the Elements API to automatically update records identified as C1s

• Can we also implement some of these automatic filters when the record is first harvested?

• Can we remove the need for researchers to complete the deposit process for every publication?
Other solutions and improvements
Faculty would view each and every publication before assigning a HERDC classification.

Library would then add metadata etc to aid discovery, make OA and publish into DRO.

A nightly feed fed metadata from DRO.

Elements has become the source of publication truth.

Metadata fed directly from Elements to Research office and staff profiles.

Then into Research Master...

... and Staff profiles.
New supporting guidelines and policies

Research Repository Procedure
- High level document approved by Academic Board in March 2017
- Defines Roles and responsibilities of Researchers, Faculty, the Library and Deakin Research
- Splits research publications into Reportable and Non reportable material

Research Guidance Sheet
- An official Research Guidance Sheet authored by Deakin Research
- Aimed at Faculty administrators
- Advocates a consistent ‘One Deakin’ approach to processing publications
- Officially sanctions the ‘Smart ingest’ principles to identify C1s

Research Repository Guideline
- Outlines the role and purpose of the repository, emphasising preservation, discovery, and making research open access
- Specifies what research outputs DRO collects (along with other material collected by the Library)
Revamped wiki space to provide assistance and guidance for researchers

Provides researchers with info about:
- The publication workflow, and roles and responsibilities
- How to optimise their profile to facilitate efficient processing
- Linking their ORCID and Elements accounts
- How to use Elements
- FAQs
- Tools and resources for download
- How to import and export
- Contact us
- Feedback
- Policies and guidelines, etc.
Toolkit of downloadable infographics, e.g. ‘How to Optimise your profile’

Follow these steps ... to streamline the processing of your publications

1. Sign the DRO Deposit Agreement
   - The agreement is between the author and the University, and enables DRO to deposit your work, which publisher will provide on your behalf and where possible, provide the reader access to open access.

2. Get an ORCID ID
   - An ORCID is a unique researcher number that works as a persistent digital ID. It distinguishes you from other researchers and helps smooth the way for the automatic matching and processing of publication metadata from one system to another.

3. Connect your ORCID and Elements accounts

4. Add your researcher IDs to your Elements search settings
   - If you have a Scopus, Web of Science or other researcher ID, you can add these to your researcher profile in Elements. If Elements finds a publication that matches your researcher ID, it will automatically claim the publication for you.

5. Add the name of your previous institution to your Elements search profile

Solutions and improvements
DeakinSync is a personal portal which provides access to various University services

The new ‘Progress of your publications’ tile provides:

- Links to any actions in Elements that are required by the researcher
- Lists of publications that are still being processed either by Faculty or by the Library
- A list of all records published to DRO

Library and Faculty staff can access this information, which helps out with troubleshooting and queries with regard to where publications are in the workflow.
Next steps

• Focus on Scopus ID cleanup and uptake

• Work with Symplectic:
  • to develop the Elements API to automatically update articles identified as C1s
  • Investigate further improvements to support automated processing

• Work with Faculty:
  • to ensure they are happy and confident with the new approach
  • to target the non-reportable records that are stuck in the queue
Questions?

michelle.watson@deakin.edu.au