Developing research data services with thesis deposit

JANET AUCOCK, UNIVERSITY OF ST ANDREWS
FEDERICA FINA, UNIVERSITY OF ST ANDREWS
Introduction
• Research intensive
• 10,000 students
• c.1150 PGR (February 2018)
• Up to 250 new theses annually
The aim of developing the services

• To create an environment where students are encouraged to deposit research data alongside their full text thesis
• Raise awareness of best practice for:
  – Students
  – Supervisors
  – Institution
The drivers

- Funders’ data sharing policies
- Funder acknowledgement requirements
- Improved management of content
- Non-textual files already being submitted. Survey showed 50+ theses with MPEG, MOV, AVI, CIF, XLSX, ZIP additions across subject disciplines
The drivers

- ETD Conference 2016
- Augmented ETDs
- Supplementary data or underpinning data
The project so far

• Started in Spring 2017
• Agreed as an objective in the Library operational plan
• Core team: Federica Fina (RDM), Janet Aucock and David Collins (E-theses, Library), Jacqui Ritchie (Registry, Continuous improvement)
• Stakeholders identified as: PGR students, supervisors, Pro-Dean Research Postgraduate, Research Data Management (RDM) team, E-theses team and Registry team
The project so far

- Methodology: review existing processes, workflows, systems and have regular working meetings
- No extra funding or staffing
- New process live on 30 January 2018
- Opportunities:
  - Review of current declaration to include research data
  - Upon approaching the Pro-Dean we discovered development work being carried out in Registry
The E-Theses service

- E-theses deposit service since 2007
- 1 FT administrator, 1x.2FT manager
- February 2018. 2775 theses in St Andrews Research Repository
- Student required to submit a thesis declaration
- Very precise embargo handling on full text, abstract and title
- 3 student training workshops annually, daily enquiries
- Current theses plus digitization programme for 3.5K theses 1919-2006
- https://www.st-andrews.ac.uk/library/resources/collections/theses/
The RDM service

• Introduced in May 2015
• 2 FTEs
• Advise on funders’ policies, data management plans, data sharing and best practice in data management
• Main funders: AHRC, BBSRC, EPSRC, ESRC, MRC, NERC, STFC, H2020
• Training to staff and postgraduate students
The institutional policies

- Data management policy – [https://www.st-andrews.ac.uk/staff/policy/research/researchdata/](https://www.st-andrews.ac.uk/staff/policy/research/researchdata/)


- No mandate for students to deposit research data but we had to obtain Pro-Dean approval for the new process and inclusion in the thesis declaration
The systems
DSpace

- St Andrews Research Repository 2007-
- 11.5k items, all full text (except embargoed theses)
- Fed by PURE CRIS for publications and direct thesis deposit
- Harvested by EThOS and Library Discovery service
- Good visibility and usage
- https://research-repository.st-andrews.ac.uk/
Pure

- Institutional Current Research Information System (CRIS)
- Introduced in 2010 and integrated with the open access institutional repository (DSpace)
- In 2015, research data catalogue and repository
- Linking across content
- Public portal for content discoverability (https://risweb.st-andrews.ac.uk/portal/)

Activities
Publications
Activities
Impact

University Structure [HR]
Staff Records [HR]
Student Records [Registry]
Projects, Grants, KT [Finance]

Research data
Entered
Entered
Entered
Fed Out

Pulled In
Linked
Linked & held

Collaboration, reporting, strategic planning etc.

Full Text Repository
Open Access
Research datasets (local storage)
Content linking and metadata across systems

**PURE data record**
- Full metadata
- Self deposit and mediated deposit
- Data DOI assigned and stored
- Links to authors, supervisors
- Links to project and funder
- Embargo handling on research data files
- Embargo handling on description and title (PG user or FF)

**PURE thesis record**
- Brief metadata
- Mediated deposit
- Links to authors, supervisors
- Links to project and funder
- Handle link to full metadata and full text
- Embargo handling on abstract and title (DC)

**DSpace thesis record**
- Full metadata and full text subject to embargo
- Self deposit and mediated deposit
- Handle assigned and stored
- Potentially DOI for thesis full text
- Will reference associated research data by DOI and descriptive note
- Embargo handling on full text, abstract and title (PG user or DC)

**Library Catalogue and Discovery Service/ETHOS**
- One way link from PURE metadata to fuller metadata and full text in DSpace. Uses DSpace handle to create the link

Future development

One way link from DSpace metadata to associated research data record(s) in PURE. Uses DOI for research data to create the link and uses dc.relation.uri to store the DOI. Link only created if DOI for data is active.

Two way link between PURE record for research data and thesis metadata
When catalysts go bad: chemistry in the classroom and the road to the liminal

Abstract
MB-remarkeable hinnone porrets were alladium convesgated. The synthesis of compounds onto t-octane was rendered alkene-resistant onto homogeneous coupling commenss to allow facile product separation of this and of successful catalyst syst-prod separation products, as well as by lactone hine peers. Resins were output successfully to become homoversal propanate. Any large-scale passing of these current modificataters was then immobilised. Tothyl product separation of this coupling commenss catalys-tics was investigated for heavier problems. Natural strawberry flavouring was added to the resulting compound, and the final refined crystals were then administered to a cohort of thirty pre-screened subjects. Members of St John's Ambulance and the Fire Brigade were on hand at all times.

Metadata
Record
Title
When catalysts go bad: chemistry in the classroom and the road to the liminal
Authors
David Collins
David Collins
Institution
University of St Andrews
Rights
Attribution-ShareAlike 4.0 International
http://creativecommons.org/licenses/by-sa/4.0/
Collections
Computer Science Theses

Keywords
Participatory design
Tablet technology
Learning using chemical technology
In-the-wild studies
Handwriting artifacts

Description of related resources
Data underpinning - Dynamic covalent assembly and disassembly of nanoparticle aggregates Barisay, S., Kay, E. R., University of St Andrews, 4 Apr 2016 DOI: 10.1783/4639d9e9-a8c3-44b4-9eb2-bbc639384a6

Related resources
http://dx.doi.org/10.1783/4639d9e9-a8c3-44b4-9eb2-bbc639384a6
http://dx.doi.org/10.5281/zenodo.34012
Declaration tool

- SITS system from TRIBAL
- Not integrated with PURE or DSpace.
- Project led by Jacqueline Ritchie (Registry)
- Allows students to build their thesis declaration by answering a series of questions
- Students can ask to place an embargo on thesis and/or data
- Self service and student facing
- Standardized output
- Step-by-step guidance document was developed
- Acknowledgements and data access statement template
The tasks in this container relate to the preparation and examination of your thesis and should only be completed when you are nearing submission of your thesis. You may, however, wish to read the documents below at any time for information.

The ‘Guidance for submission of theses’ document provides a step-by-step guide to completing your tasks and it is essential that you read it prior to the submission process. It will be particularly helpful to you when discussing submission with your supervisor(s), especially in relation to the decisions you will need to make in your declarations. It also provides information on the presentation of your thesis.

The ‘Research degree examinations’ document outlines the research degree examination process and is aimed at research degree students, supervisors, and examining committees.

Guidance documents
- Guidance for submission of theses (PDF)
- Research degree examinations (PDF)

Template(s)
- Acknowledgement page (DOCX)

Preparing for submission of my thesis
**Electronic copy of your thesis**

These are the same basic questions as for the Printed copy of your thesis.

**Q. Would you like to embargo the electronic copy of your thesis?**
This is a ‘yes’, ‘no’ answer. If you answer ‘no’, you move on to the ‘Permission for publication of underpinning research data/digital outputs’ section. If you answer ‘yes’, you are asked the following questions:

**Q. Would you like to embargo all or part of the electronic copy of your thesis?**
You can select either ‘all’ or ‘part’.

Note that ‘all’ or ‘part’ does not include the title and abstract. If you choose to embargo, you will be asked separately if you would like to embargo your title and/or abstract of the full text of your thesis.

---

**Copyright**

**Q. Does your thesis contain any third party material that requires copyright clearance?**
This is a ‘yes’, ‘no’ answer. Depending on your answer, it displays some relevant information. More general information can be found in the box below.

**Useful background information about copyright**

“Your print thesis is an unpublished work and has historically been viewed as an examination script. Copyright for the purpose of examination is now provided for under ‘Illustration for Instruction’ (CIPA s.22, June 2014) and so the legislation allows you to include third party copyright material provided it is accompanied by sufficient acknowledgement and its use is considered ‘fair dealing’, i.e. the amount of the work being copied is reasonable and proportionate for your purposes.

Your digital (electronic) thesis will be made available online, and is therefore deemed to be published so you will have different responsibilities with regard to any third party copyright material you use, e.g. quotations and extracts from other people’s publications, i.e. books and journals, illustrations, photographs, diagrams, maps, graphs. If you are unable to obtain permissions then you may need to consider embargoring part of your electronic thesis.” Extract from the GRADskills Microsoft PowerPoint presentation, ‘Copyright 101 November 2016’, available on the Library’s Thesis webpage.

---

**Acknowledgement of funders**

**Q. Did you receive funding from an organisation or institution?**
This is a ‘yes’, ‘no’ answer. Depending on your answer, it displays a note of what will be stated in your declarations.

**Useful background information about funders**

Most major research funders include in the terms and conditions of their grant awards a requirement for students to acknowledge in their theses the support received from the funder in question.

---

**Underpinning research data or digital outputs**

**Q. Is your thesis based on original research data/digital outputs that are subject to the requirements of the University or the funding body to deposit them?**
This is a ‘yes’, ‘no’ answer. If you answer ‘no’, you move on to the ‘Confirmation’ section where you can download/save/print both your title page and declarations. If you answer ‘yes’, you are asked the following question:

**Q. Would you like to embargo your research data/digital outputs?**
This is a ‘yes’, ‘no’ answer. If you answer ‘no’, you move onto the ‘Confirmation’ section where you can download/save/print both your title page and declarations. If you answer ‘yes’, you are asked the following questions:

**Q. Would you like to embargo all or part of your research data/digital outputs?**
You can select either ‘all’ or ‘part’.

Note that ‘all’ or ‘part’ does not include the title and description. If you choose to embargo, you will be asked separately if you would like to embargo your title and/or description of your underpinning research data/digital outputs.

If you choose ‘part’, you are asked to answer the following question:

**Q. What part(s) of your thesis would you like to embargo?**
This is a free text box. You should enter the part(s) separated by commas e.g. data files 010, 065, 090, folder 1. Do not list one under another.

---

**Permission for publication of underpinning research data or digital outputs**

The University and all major research funders have requirements on the deposit and sharing of research data or research digital outputs that are created or collected as part of any research project, including data or digital objects underpinning your thesis. You should, therefore, check if specific funders’ requirements apply to your work and what you can do to meet them.

If your research is subject to requirements to deposit data/digital outputs then, in addition to deciding whether or not you need to request an embargo on the full text of your thesis, you would also have to decide whether or not there is a need to embargo the data/digital outputs.

As for the full text, the matter of an embargo on data or digital outputs should be discussed with your supervisor(s) as early as possible. Similarly to the full text, you can request an embargo on the grounds that publication would:
- be commercially damaging to the researcher, or to the supervisor, or to the University, (for example, intellectual property rights)
- preclude future publication
- be in breach of law or ethics or data protection.

Depository of your research data/digital output is via PURE, the institutional research information system and can be deposited before or after submission of your thesis for examination.

For advice and assistance, please contact the Research Data Management Team (research-data@st-andrews.ac.uk) or visit the research data management website.
General acknowledgements

Please insert your general acknowledgements here.

Funding

Please list the names of any organisations or institutions who have funded your thesis research using the specific format and punctuation below. [An example of the formatting can also be found in Appendix II of the Guidance for submission of theses document].

This work was supported by the insert funder name in full here [grant number xxxx].

OR

This work was supported by the Funder 1 Name [grant number xxxx]; and the Funder 2 Name [grant number yyyy].

OR

This work was supported by the Funder 1 Name [grant number xxxx]; the Funder 2 Name [grant number yyyy]; and the Funder 3 Name [grant number zzzz].

OR

This work was supported by the Funder Name [grant numbers xxxx, yyyy].

Research Data/Digital Outputs access statement

Research data underpinning this thesis are available at [DOI]

OR

Digital outputs underpinning this thesis are available at [DOI]
The workflow
4 months before end date: Registry invites student to complete the ‘Intention to submit’ task

Student receives RDM email and creates dataset record in Pure (metadata only)

RDM team check record

Embargo?

Yes

Emargo approved

No

Student submits for examination

Student completes task and specify 4 weeks window

Student completes ‘Preparing for submission or resubmission of your thesis’ task

RDM team receive receipt email

RDM team email data guidance to student

RDM team receive EMAIL copy of embargo letter

RDM team apply embargo to metadata (if applicable)
Examination and revisions approved

Student uploads data files in Pure and notifies RDM team

Student receives DOI and adds it to full text

Degree awarded

RDM team check files and apply embargo (if applicable)

RDM team issue inactive DOI and notify student and supervisor
Student

Final examined version submitted to registry

E-Theses

E-Theses team receive receipt email and hard copy of theses

E-Theses team checks declaration vs embargo letter

RDM

Contact Registry and cc RDM

Record in SAULCAT and Pure created and e-copy uploaded to server

Student registers to repository

Invite student to register to repository

E-Theses

RDM team receive receipt email

RDM team see e-Copy

RDM team check declaration vs embargo letter (if applicable)

RDM team check declaration for data statement

RDM team receive receipt email

6 possible cases
6 possible cases

Case 1
- Data record in Pure
- Data files in Pure

Embargo:
- No embargo
- Files only
- Files + description
- Files + description + title

Case 2
Secondary data

Case 3
Sensitive data

Case 4
- Data record in Pure
- NO Data files in Pure

Case 5
- NO Data record in Pure
- NO Data files in Pure

Case 6
- Extra files with Full Text
Lessons learnt
Lessons learnt

• Give yourself plenty of time
• Don’t overthink it – you’ll never cover all the possible outcomes anyway
• More user engagement
• Have face-to-face with relevant stakeholders
• We have been very lucky to have enthusiastic colleagues
Going forward
Going forward

- Upskilling the students and the supervisors
- Upskilling ourselves
- Monitor enquiries and feedback to improve the process
  - First enquiry: highly sensitive/confidential data, how to proceed?
  - “I do not want to be dealing with this in the four months countdown to submission.”
  - We realized we need to change the guidance document
Going forward

• Further explore the use of Pure for metadata-only theses records
• Integrate DOIs for theses
• Integrate ORCID
• Build and refine funder acknowledgement workflows to record funder metadata
• Clarify the distinction between underpinning and supplementary data
Wish list
Wish list

• More integration between systems
• Forward thinking students
• More engagement with students and supervisors
Acknowledgements
Acknowledgements

• Jacqui Ritchie - Registry Co-ordinator (Continuous Improvement), Academic Registry
• David Collins - Principal Library Assistant (Cataloguing), Library
• Liliana Martins e Caneco- Registry Officer
• Registry team