

Tentative steps in mining UK theses

OR 2016, Dublin

June 2016

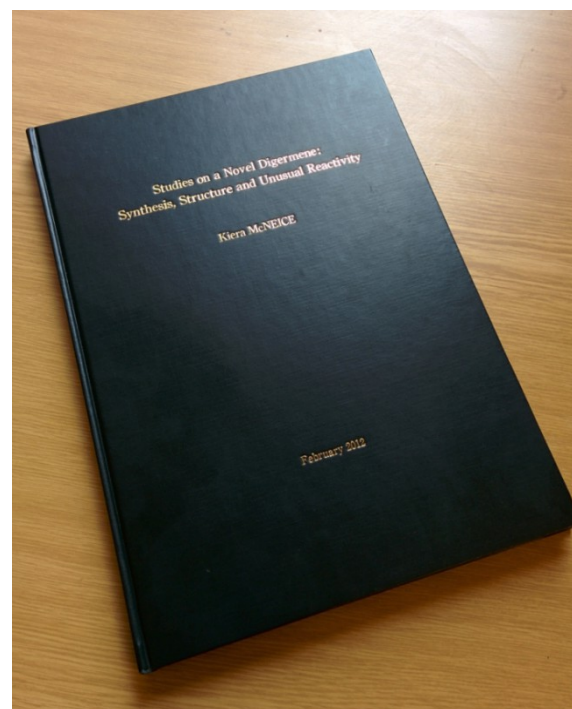
Is there valuable content in theses?

“Anything worthwhile in a thesis would have been published separately anyway.”

-- bioscience researcher

UK PhD theses

- Cutting edge research
- Not published elsewhere
- Traditionally book, now usually e-
- PDF – but new forms emerging
- 20,000 / year
- 300 pages each
- 6m pages of unique research every year



EThOS – e-theses online service



[Basic search](#) | [Advanced search](#) | [Search results](#)

[Login / Register](#) | [About](#) | [Help](#) | [FAQ](#) | [Follow](#)

Search EThOS

Search over 400,000 doctoral theses. Download instantly for your research, or order a scanned copy quickly and easily.

Limit search to items available for immediate download from EThOS or the Institution:

Can't find the thesis you want? Click [here](#) to ask our experts



"A wealth of data is buried in theses which can shed light on very interesting areas..."

Find out how three researchers from the Universities of Manchester, Plymouth and Glasgow have used EThOS to help their research.

[Find out more](#)

More about EThOS:

- ▶ [Participating institutions](#)
- ▶ [Resources in the EThOS Toolkit](#)
- ▶ [OAI-PMH harvesting](#)

Other Services from the British Library:

- ▶ [Document supply website](#)
- ▶ [British Library blogs](#)
- ▶ [Sign up to our newsletters](#)

Basic search | Advanced search | Search results

[Login / Register](#) | [About](#) | [Help](#) | [FAQ](#) | [Follow](#)

Use this URL to cite or link to this record in ETHOS: <http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.518072>



Title:	The figure of the child in the novels of Thomas Hardy
Author:	Pearce, Jessica Louise
Awarding Body:	University of Exeter
Current Institution:	University of Exeter
Date of Award:	2010

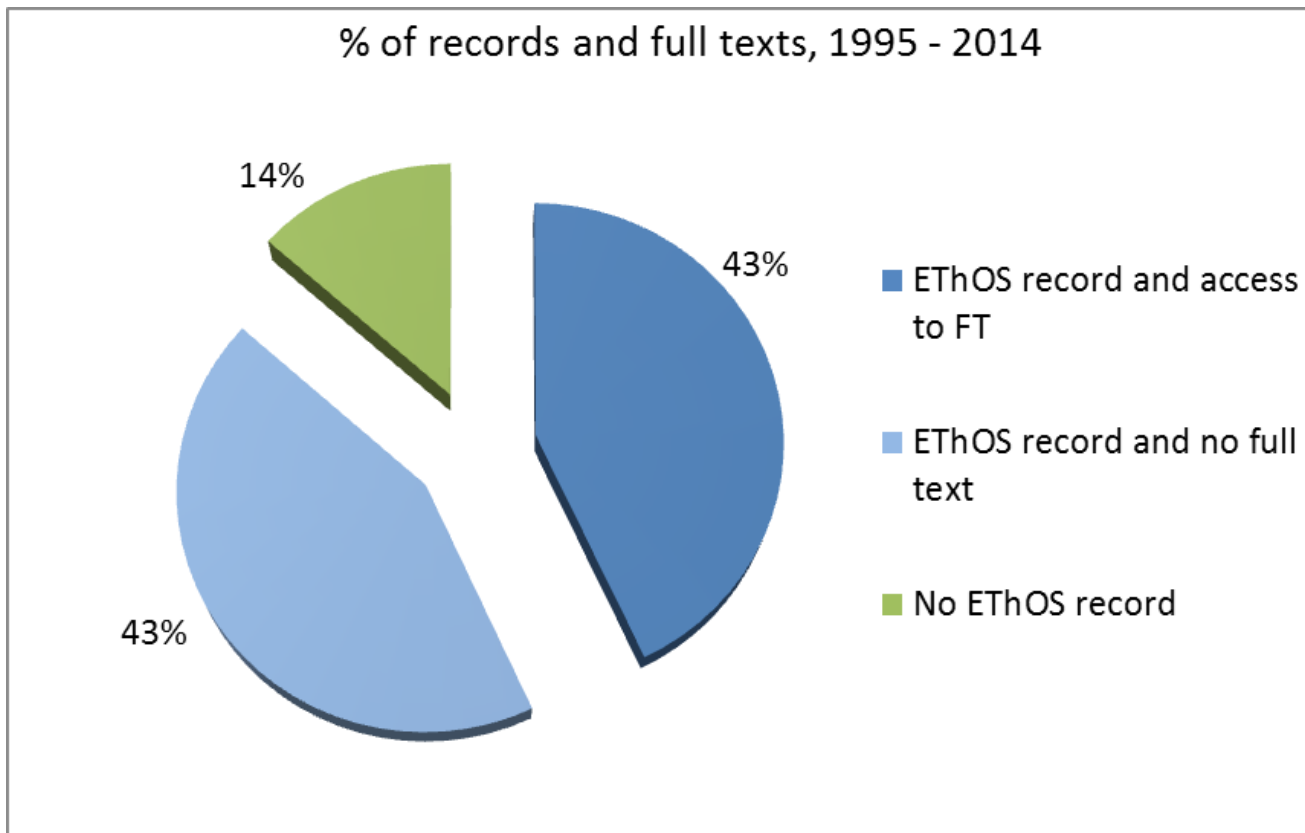
Availability of Full Text:	Access through ETHOS:	Thesis available for immediate download. Please login/register to view download & delivery options.
	Access through Institution:	http://hdl.handle.net/10036/106573

Abstract:

This study looks at the figure of the child in the novels of Thomas Hardy. It argues that Hardy, in his various presentations of the child, draws on mythologies generated by the figure of the child in the nineteenth century. The introduction describes the existence and proliferation of these mythologies during the time in which Hardy was writing. It summarizes representations of the child in history, science and literature, and reviews existing critical literature on the topic. The study comprises six chapters. The first looks at babies and young children, the second at Jude the Obscure, the third at Tess of the D'Urbervilles, the fourth at pregnant women and the fifth at the child within the family unit. The final chapter uses close reading to provide a re-evaluation of two of Hardy's 'minor' novels. Each chapter draws on specific historical contexts to reveal different aspects of the child myth. The study as a whole looks at the different ways in which Hardy uses the myth. At times he participates in it, or appears to, while at others he exposes it, or employs it to expose class and gender divisions in nineteenth century society. Ultimately, Hardy acknowledges the power of the child myth in literature and in society, while simultaneously recognising it as a fallacy that is both inaccurate and dangerous.

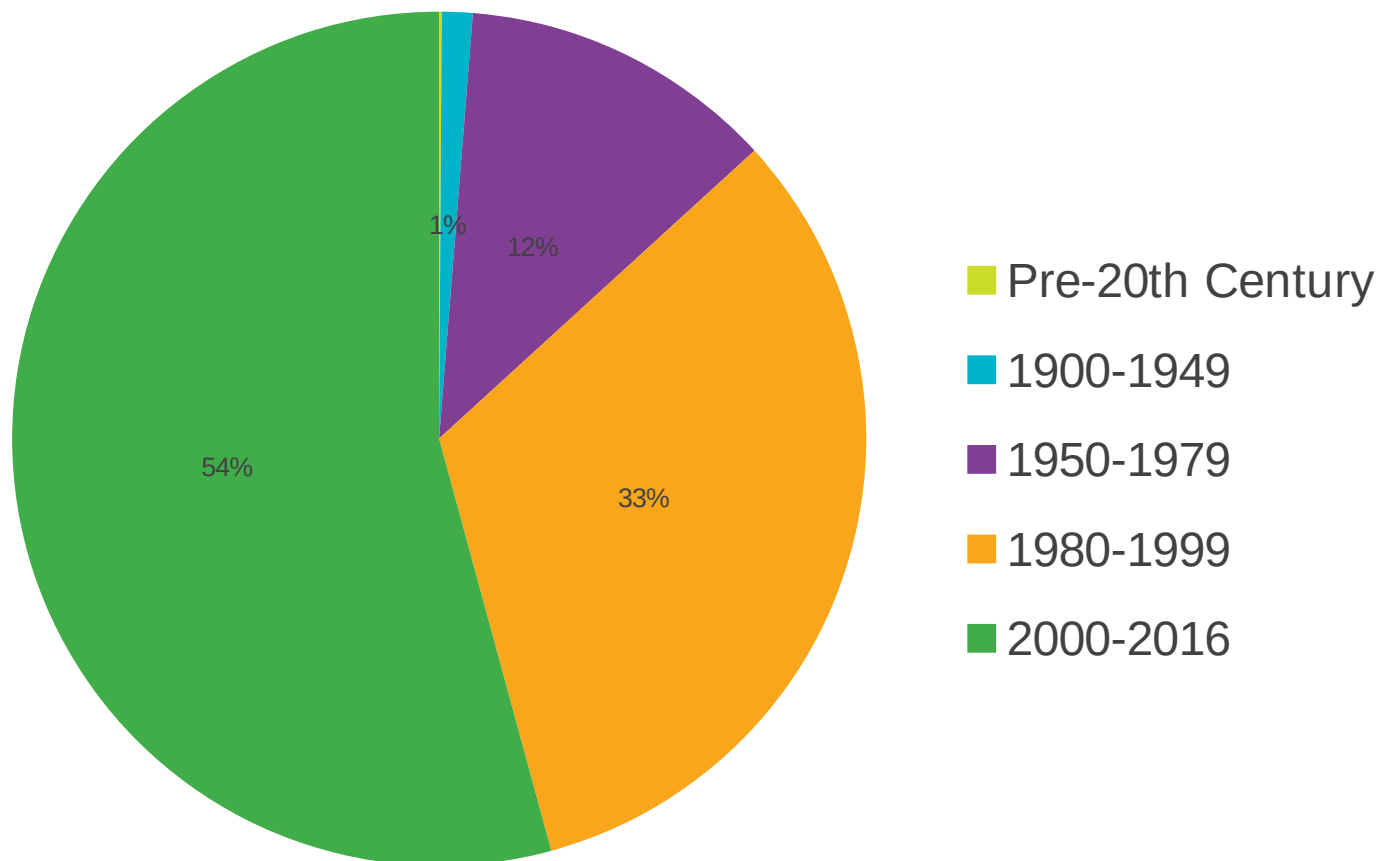
Supervisor:	Pearson, Maeve ; Richardson, Angelique	Sponsor:	Not available
Qualification Name:	Thesis (Ph.D.)	Qualification Level:	Doctoral
ETHOS ID:	uk.bl.ethos.518072	DOI:	Not available
Keywords:	Hardy, Thomas ; Child		
Share:	ShareThis Facebook Tweet LinkedIn Email CiteULike Blogger		

UK thesis collection & EThOS

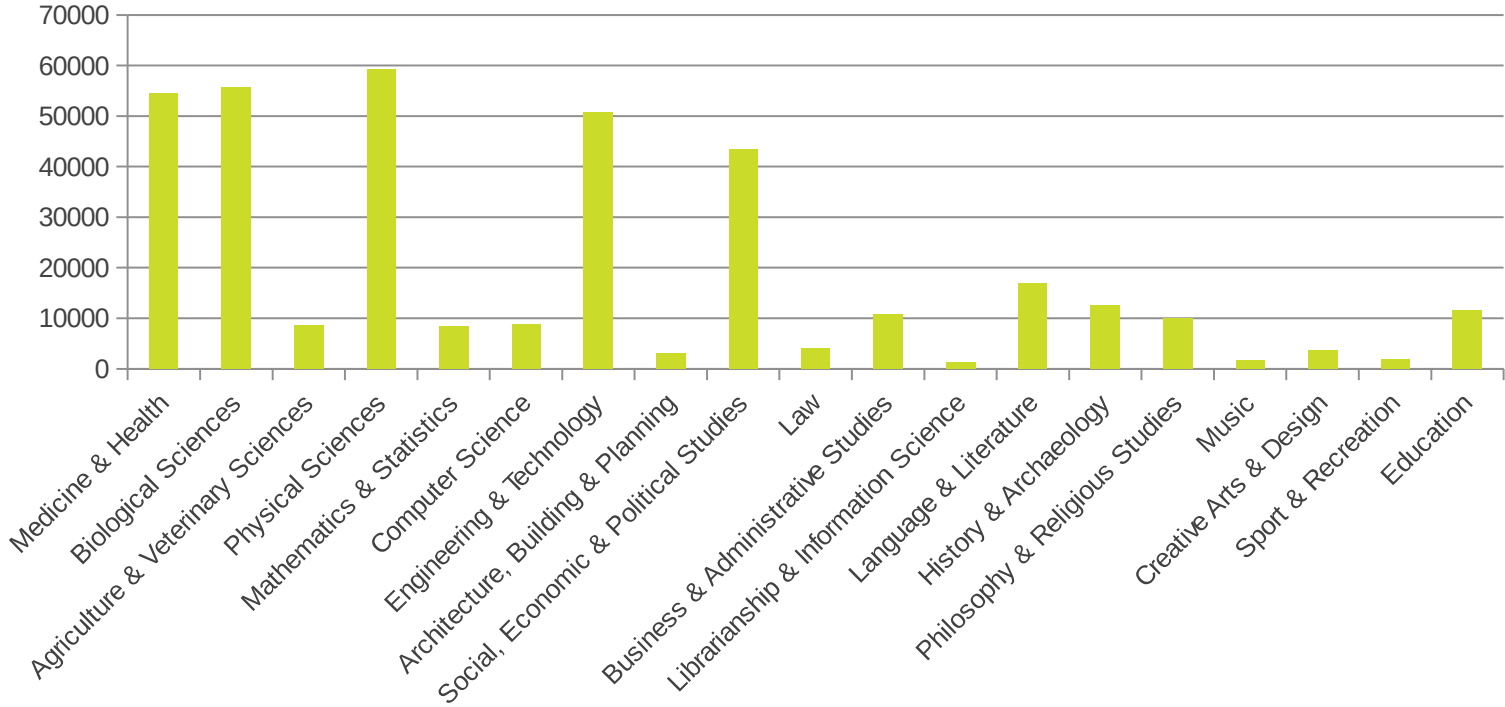


<http://ethos.bl.uk>

Theses by Date

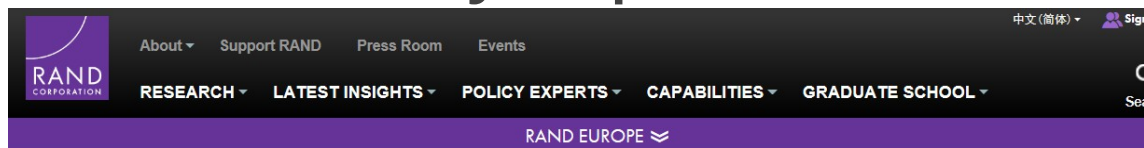


Theses by Subject



TDM examples

Alzheimer's Society report



RAND > RAND Europe > Research > Projects >

Mapping the UK's Dementia Research Landscape and Workforce Capacity

Background

According to a recent report from the Global Observatory for Ageing and Dementia Care, the number of people with dementia worldwide will rise from 36 million in 2010 to 115m in 2050. In the UK, the Alzheimer's Society has estimated that one in 14 people over age 65 currently have dementia.

Several governments, including the UK government, have announced commitments to increase funding for dementia research and innovation, with a goal to find a treatment or cure, and to improve the management of the condition through the way health and social care is delivered. It is vital that this funding be allocated effectively, building on existing strengths and targeting research gaps.



Illustration by Stephen Magrath, Wellcome Images/CC BY-NC-ND 4.0

Project Team

Sonja Marjanovic
Research Leader



Sonja Marjanovic is a research leader at RAND Europe. Her work

focuses on health research system strengthening, multi-stakeholder collaboration in the health sector, capacity building, open-innovation and dementia innovation systems. Recent projects include studies and evaluations of R&D..

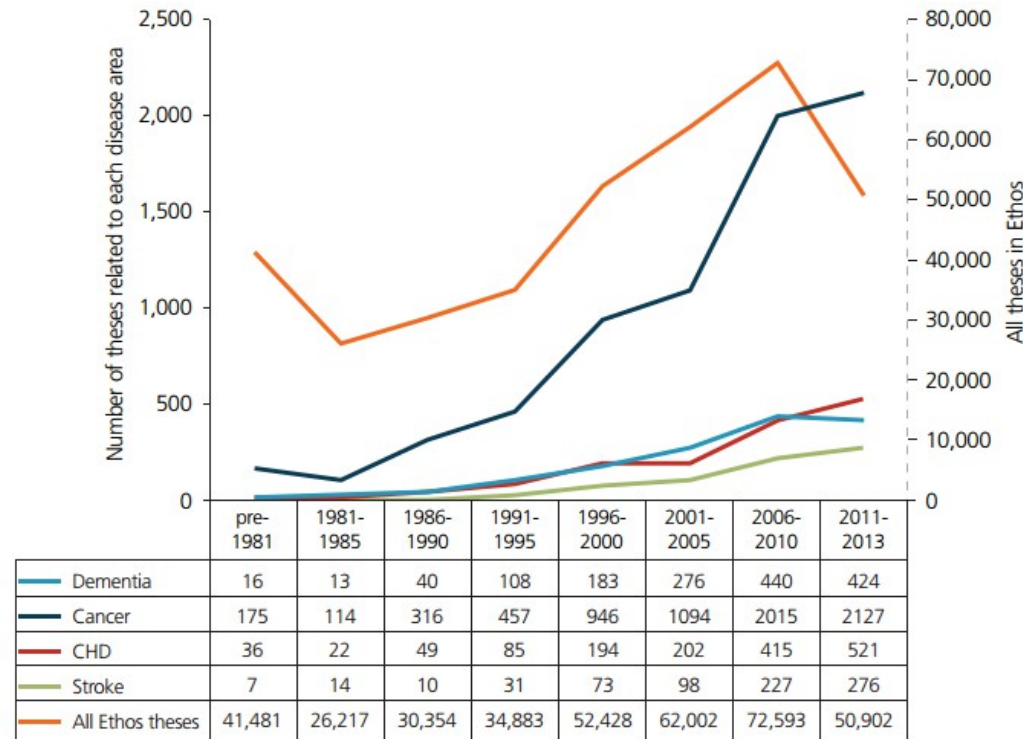
<http://www.rand.org/randeurope/research/projects/mapping-uk-dementia-research-landscape.html>

TDM case study - Alzheimer's Society & RAND Europe

Mapping the UK's Dementia Research Landscape

- Workforce pipeline
- Tracked PhD to senior research
- 1/5 dementia PhD graduates remain in dementia research
- 70% leave dementia research within 4 years of completing PhD
- Used EThOS metadata to analyse trends

Number of theses related to dementia, cancer, CHD and stroke.

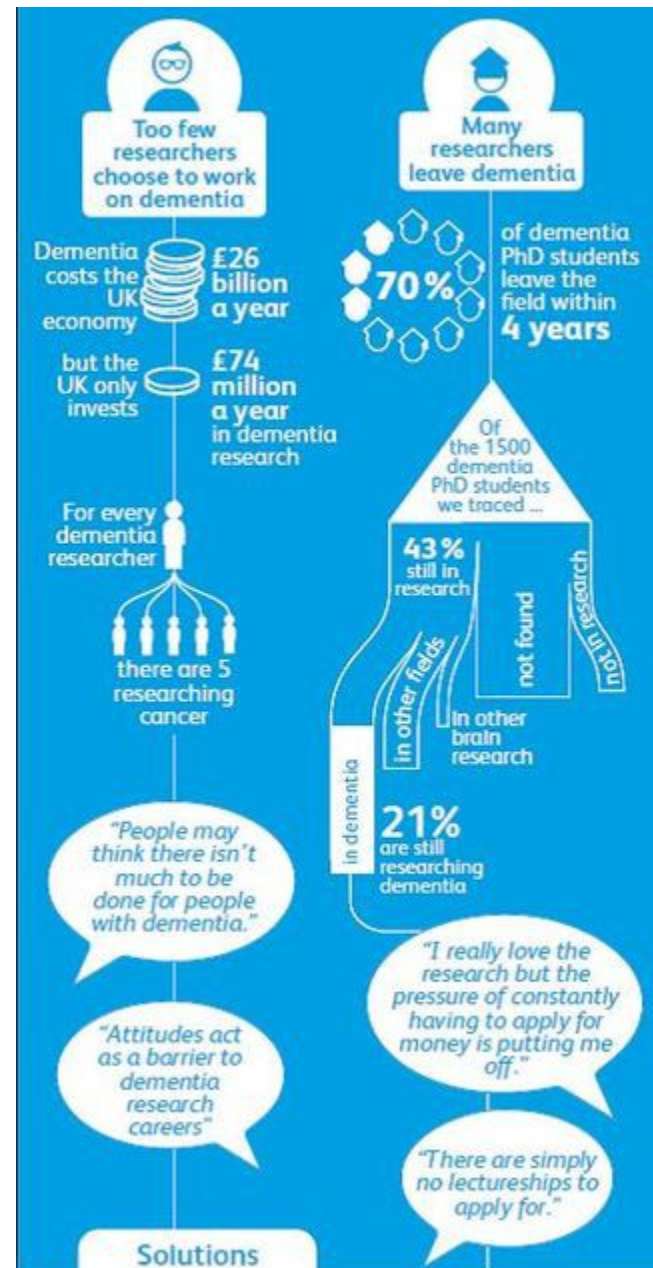


source: EThOS at the British Library

<http://britishlibrary.typepad.co.uk/science/2015/09/a-novel-use-of-phd-data.html>

Dementia search terms

- Alzheimer's
- Dementia
- Cognitive impairment
- Mixed dementia
- Early onset dementia
- Vascular dementia
- Lewy bodies (Dementia with Lewy bodies)
- Frontotemporal dementia
- Posterior Cortical Atrophy
- Familial dementia
- Creutzfeldt Jakob
- Korsakoff's syndrome
- Cognitive impairment
- Supranuclear palsy
- Binswanger's
- Multiple sclerosis
- Motor neurone disease
- Parkinson's
- Huntington's



FLAX Interactive Language Learning

flax
interactive language learning

library demos downloads about

Law PhD Thesis Abstracts

About Collection Search Browse Activities Collocations Wordlist My Cherry Basket

Browse Collocations in Collection

a b c d e f g h i j k l m n o p q r s t u v w y Top 100

Top 100 collocations

Noun+Noun (100) Adjective+Noun (100) Noun+of+Noun (100) Verb+Noun (100) Verb+Preposition+Noun (100) Adjective+to+Verb (22) Adjective+Preposition+Noun (37)

- [tax law](#) (10)
- [identity theft](#) (6)
- [tax avoidance](#) (6)
- [stem cell lines](#) (6)
- [documentary credits](#) (5)
- [thesis analyses](#) (4)
- [interface norms](#) (4)
- [jury independence](#) (4)
- [file sharing](#) (4)
- [rights law](#) (4)
- [negligence manslaughter](#) (4)

- <http://flax.nzdl.org/greenstone3/flax?a=fp&sa=library>
- Article - <http://www.journals.elsevier.com/learning-culture-and-social-interaction/>

TDM case study – FLAX interactive language learning

- Model writing at research level; domain-specific texts; co-located phrases
- Auto extraction & re-use for language learning
- Used EThOS metadata abstracts
- University of Waikato & Queen Mary, London



[library](#) [demos](#) [downloads](#) [about](#)

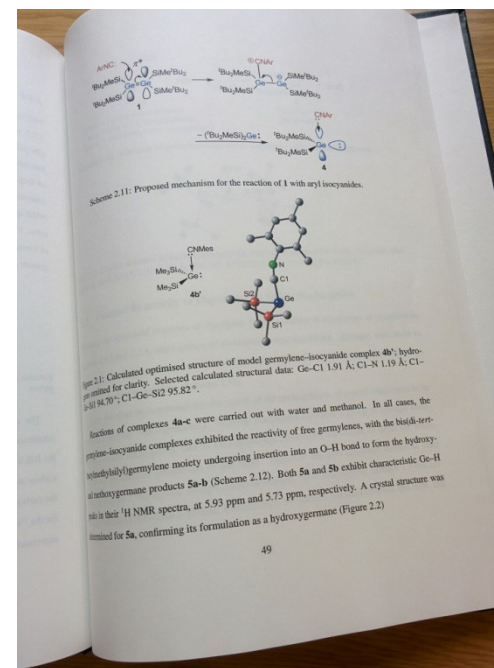
Law PhD Thesis Abstracts

Metadata or full text theses?

	Metadata	Full texts
<i>Content</i>	400,000 records	130,000 theses
<i>Format</i>	Data	<ul style="list-style-type: none"> - Digitised from print - E-born
<i>File format</i>	Xml or Excel	PDF, .wav, .mov ...
<i>Access</i>	<ul style="list-style-type: none"> - Harvest via OAI-PMH - Supplied data 	<ul style="list-style-type: none"> - Download from EThOS or other repository - Supplied with permissions
<i>Rights</i>	In the public domain	Rights holders

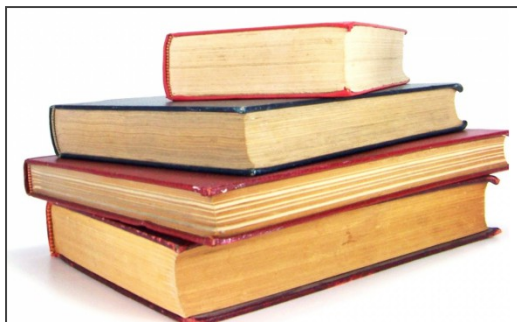
TDM case study – National Compound Collection

- Are there useful molecules in PhD theses?
- Extract the compounds; re-draw in ChemDraw; input into ChemSpider
- Bristol Uni & Royal Society Chemistry
- Manual pilot – could process be automated?
- Used theses “likely to reveal new compounds”
- 47k compounds discovered (50% new)

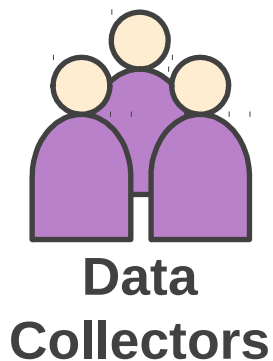


Data collection

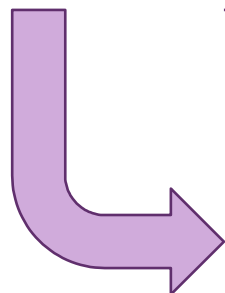
Theses



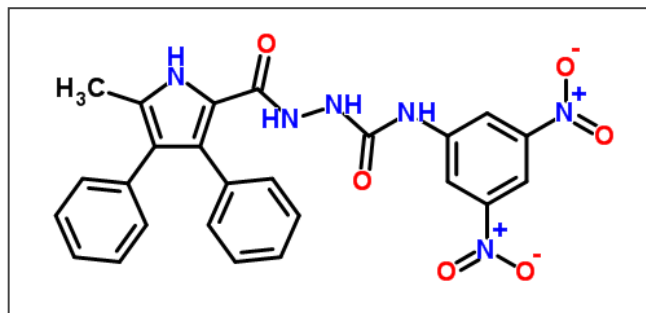
Open Access Database



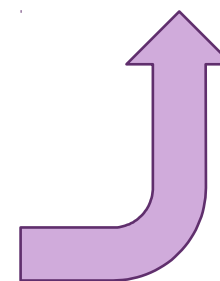
Data
Collectors



Molecular Structures



N-(3,5-Dinitrophenyl)-2-[(5-methyl-3,4-diphenyl-1H-pyrrol-2-yl)carbonyl]hydrazinecarboxamide
Louise Sarah Evans, University of Southampton, 2006



> 45,000
compounds

ETHOS – <http://ethos.bl.uk>

- Metadata for all UK doctoral (PhD) theses
- 430,000 records
- Top quality, accurate, consistent, unduplicated metadata
- Unique research, often not published elsewhere, cutting edge
- Data includes:
 - Author, title, year, university name
 - Abstracts (for 40%)
 - Supervisor names, funder/sponsor body
 - A few DOI and ORCID identifiers
 - Subject discipline.

Summary - EThOS data available

- Excel or XML via OAI-PMH harvest:

<http://simba.cs.uct.ac.za/~>

<ethos/cgi-bin/OAI-XMLFile-2.21/XMLFile/ethos/oai.pl>

- Data.bl.uk (coming soon)

Thank you

Sara.Gould@bl.uk
ethos.bl.uk