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

EThOS
e-theses online service

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

UNIVERSITY OF
EXETER

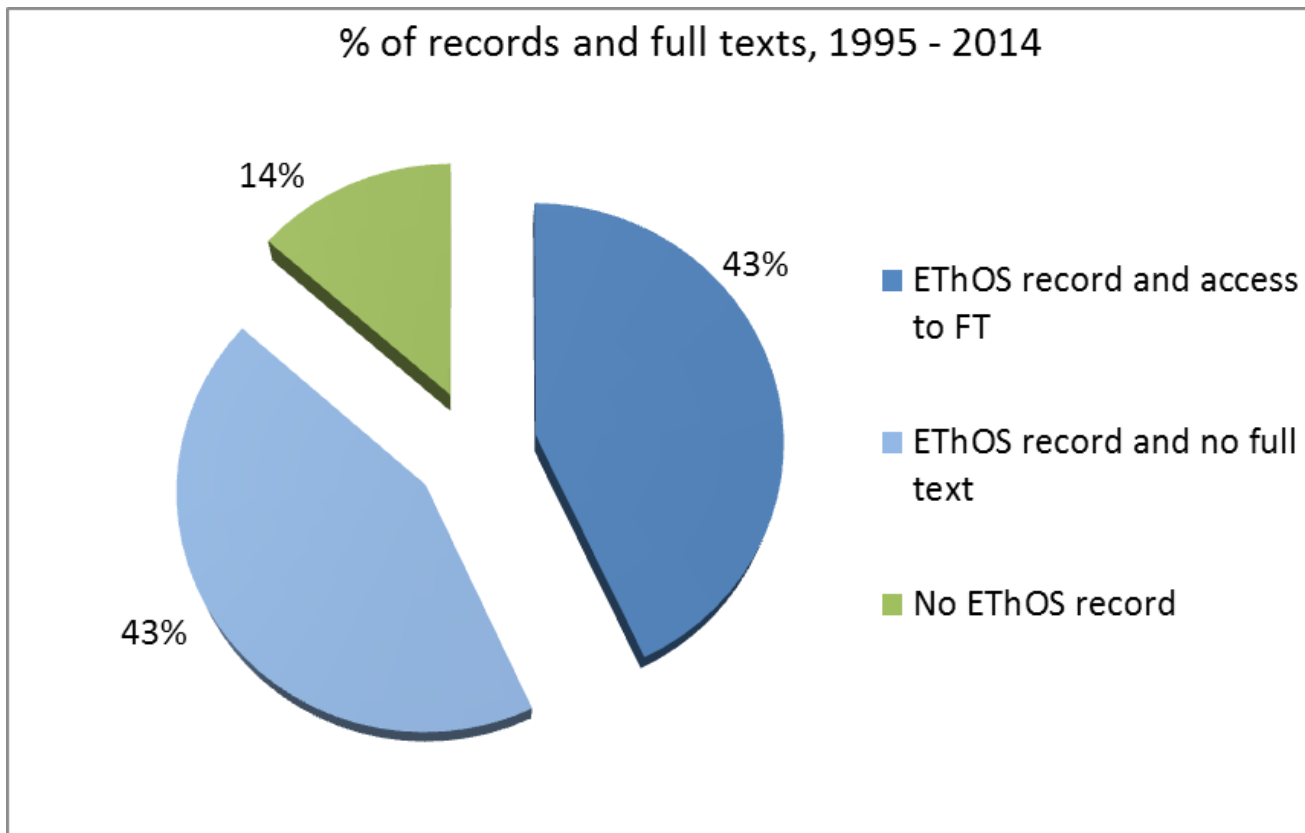| Title:	The figure of the child in the novels of Thomas Hardy
Author:	Pearce, Jessica Louise
Awarding Body:	University of Exeter
Current Institution:	University of Exeter
Date of Award:	2010
Availability of Full Text:	Access through EThOS: Thesis available for immediate download. Please login/register 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.
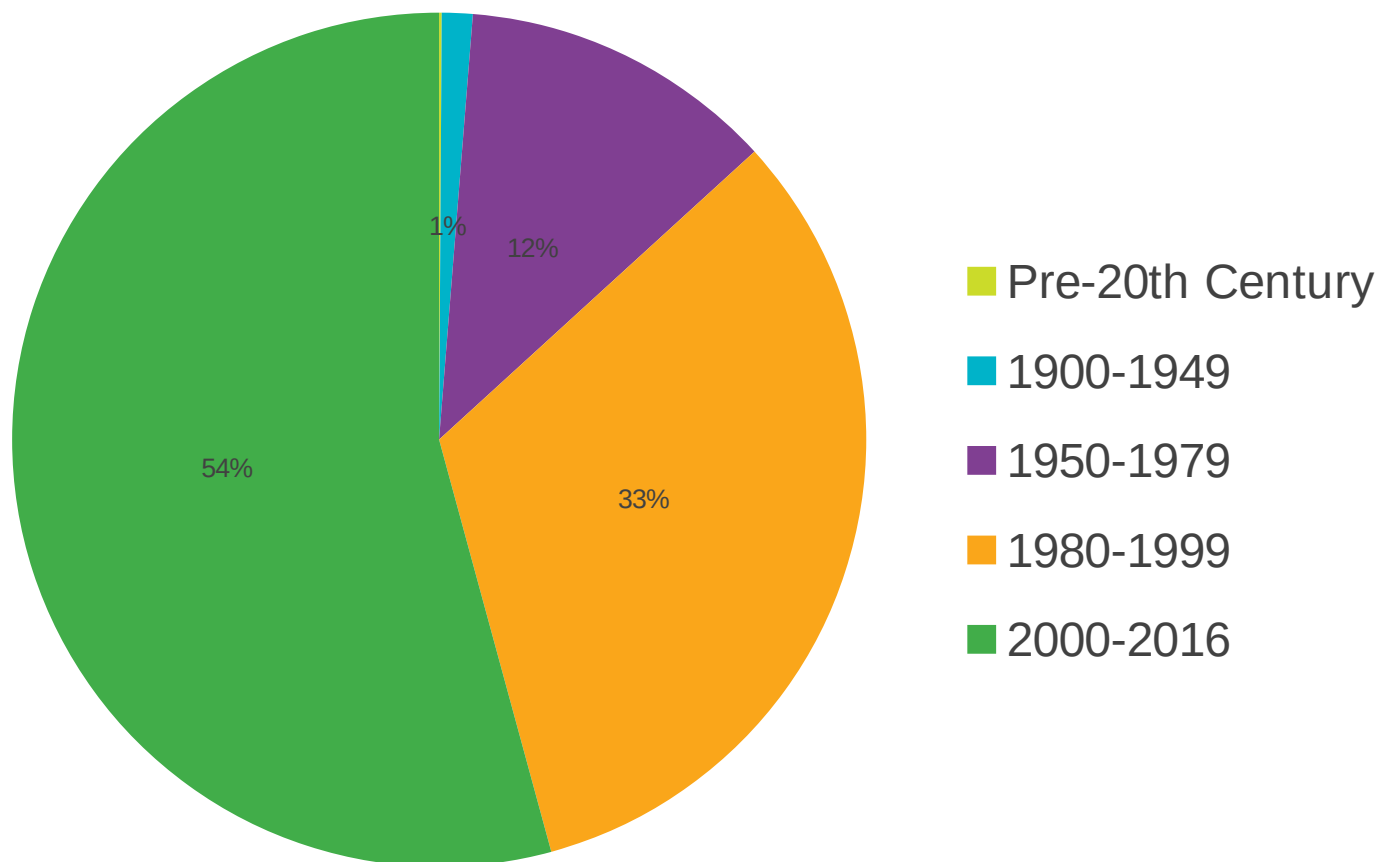
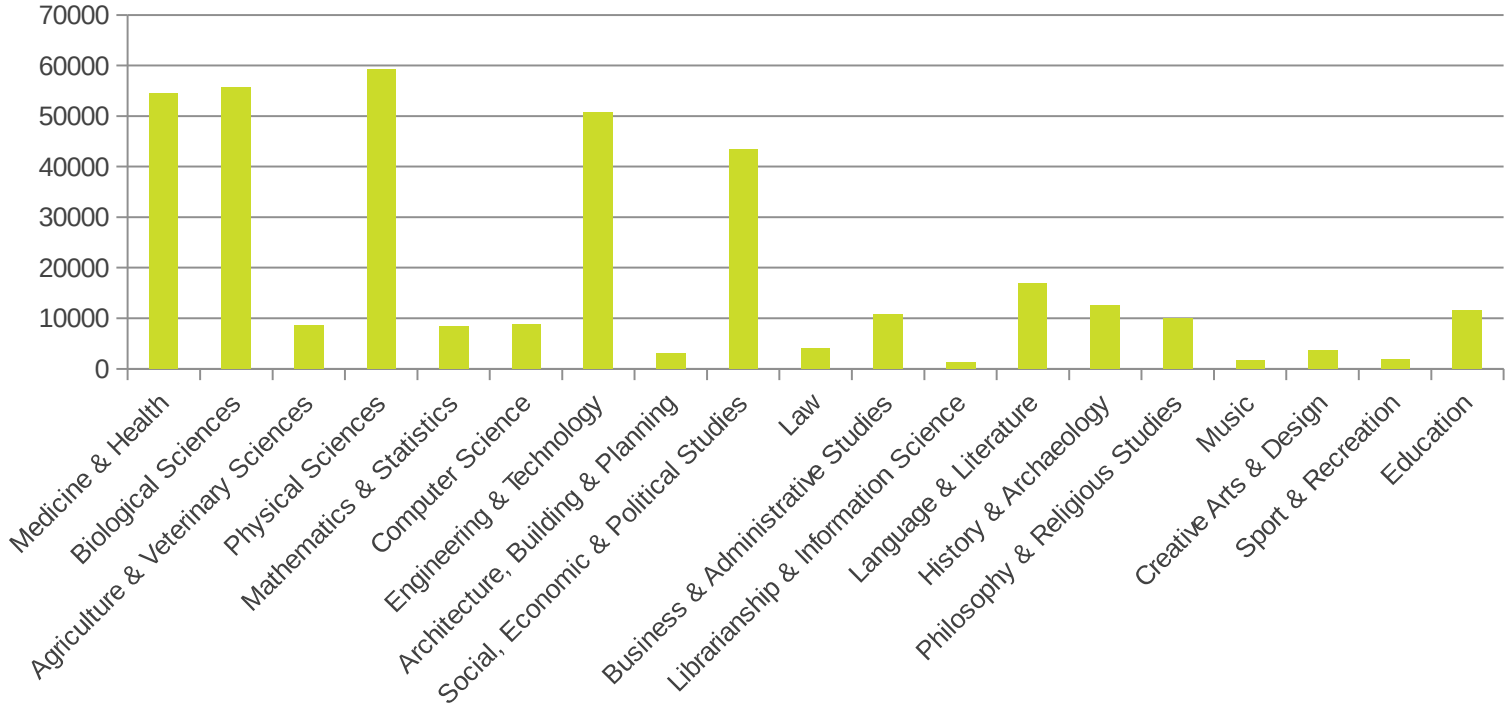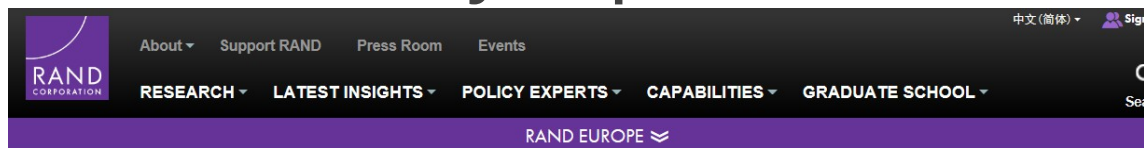| Supervisor:	Pearson, Maeve ; Richardson, Angelique	Sponsor:	Not available
Qualification Name:	Thesis (Ph.D.)	Qualification Level:	Doctoral
EThOS ID:	uk.bl.ethos.518072	DOI:	Not available
Keywords:	Hardy, Thomas ; Child		
Share:	ShareThis Facebook Tweet LinkedIn Email CiteULike Blogger		

UK thesis collection & EThOS

% of records and full texts, 1995 - 2014

- EThOS record and access to FT — 43%
- EThOS record and no full text — 43%
- No EThOS record — 14%

http://ethos.bl.uk

Theses by Date

Pie chart legend:
- Pre-20th Century
- 1900-1949
- 1950-1979
- 1980-1999
- 2000-2016

Pie chart values: 1%, 12%, 33%, 54%

Theses by Subject

TDM examples

Alzheimer's Society report

http://
www.rand.org/randeurope/research/projects/mapping-uk-dem
entia-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

ementia, cancer, CHD and stroke.

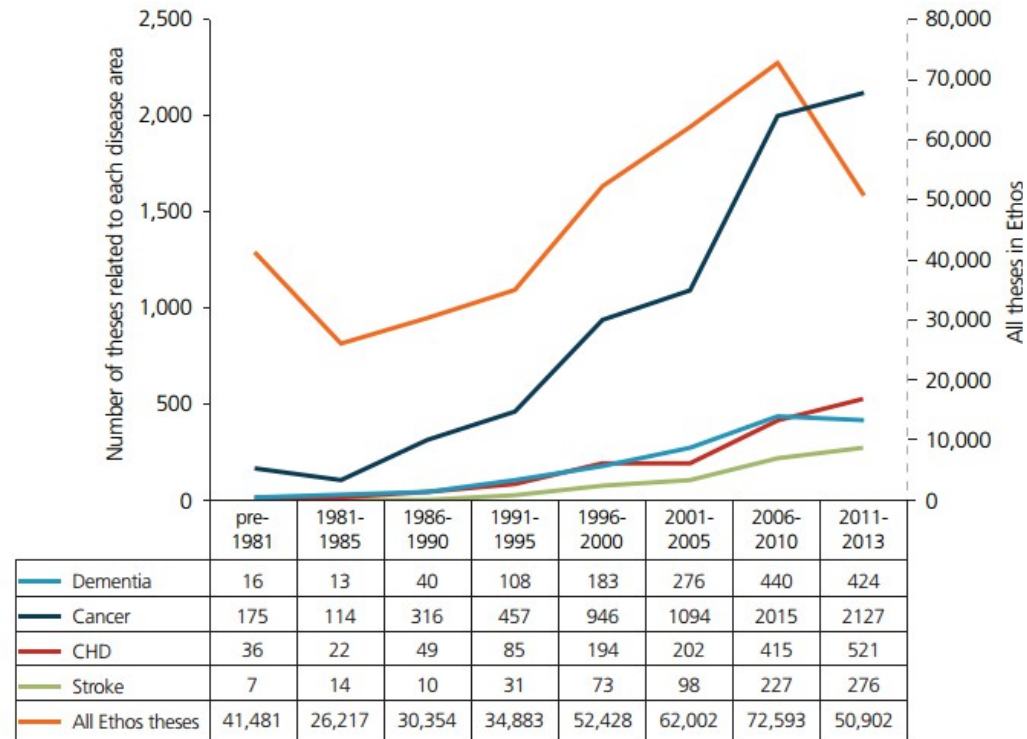|	pre-1981	1981-1985	1986-1990	1991-1995	1996-2000	2001-2005	2006-2010	2011-2013
Dementia	16	13	40	108	183	276	440	424
Cancer	175	114	316	457	946	1094	2015	2127
CHD	36	22	49	85	194	202	415	521
Stroke	7	14	10	31	73	98	227	276
All Ethos theses	41,481	26,217	30,354	34,883	52,428	62,002	72,593	50,902

Source: EThOS at the British Library.

http://britishlibrary.typepad.co.uk/science/2015/09/a-novel-use-of-phd-data.html
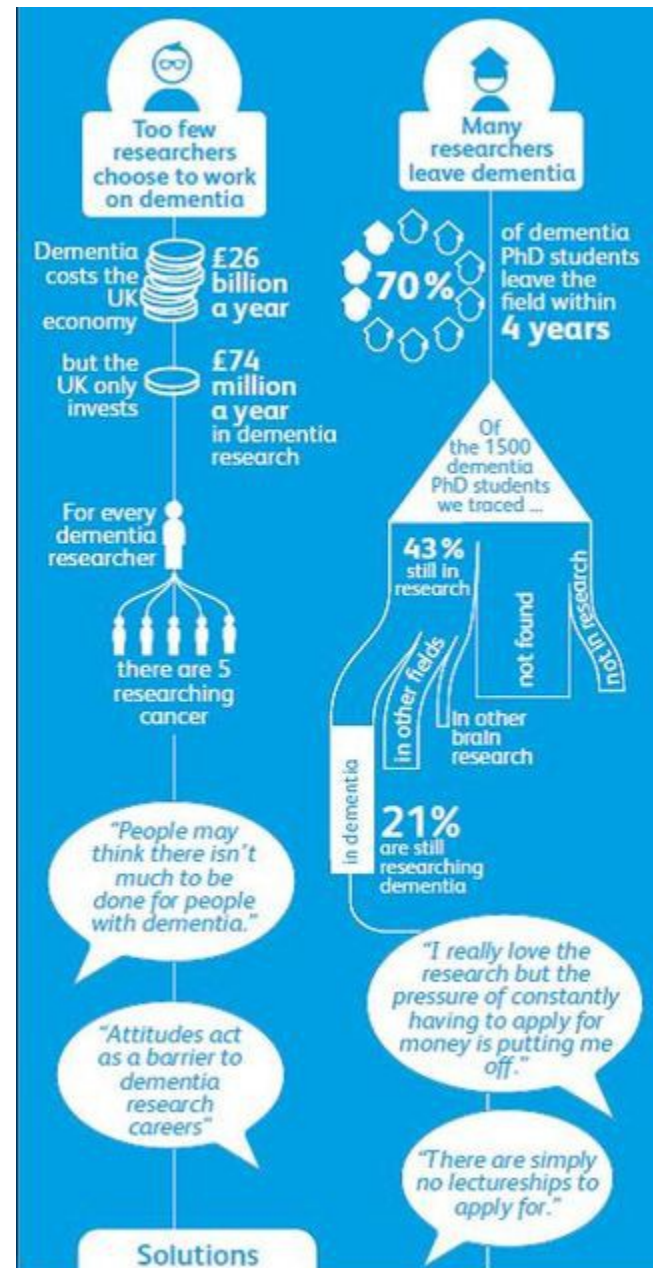
Dementia search terms

- 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

Too few researchers choose to work on dementia

Dementia costs the UK economy £26 billion a year

but the UK only invests £74 million a year in dementia research

For every dementia researcher there are 5 researching cancer

"People may think there isn't much to be done for people with dementia."

"Attitudes act as a barrier to dementia research careers"

Many researchers leave dementia

70% of dementia PhD students leave the field within 4 years

Of the 1500 dementia PhD students we traced ...

43% still in research

21% are still researching dementia

in other fields · not found · not in research · in other brain research · in dementia

"I really love the research but the pressure of constantly having to apply for money is putting me off."

"There are simply no lectureships to apply for."

Solutions

FLAX Interactive Language Learning

- 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

flax
interactive language learning

library   demos   downloads   about

Law PhD Thesis Abstracts

Metadata or full text theses?

|	Metadata	Full texts
Content	400,000 records	130,000 theses
Format	Data	- Digitised from print - E-born
File format	Xml or Excel	PDF, .wav, .mov …
Access	- Harvest via OAI-PMH - Supplied data	- Download from EThOS or other repository - Supplied with permissions
Rights	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"
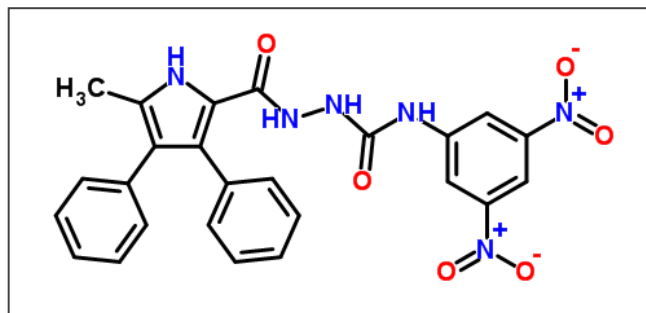
- 47k compounds discovered (50% new)
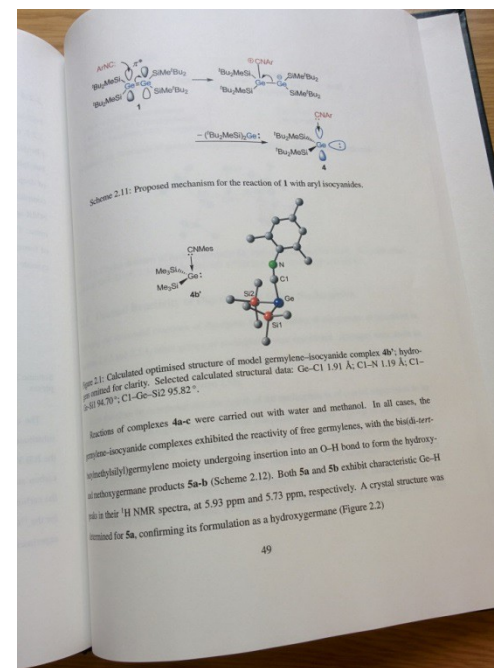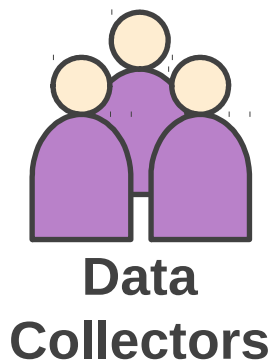
Data collection

Theses
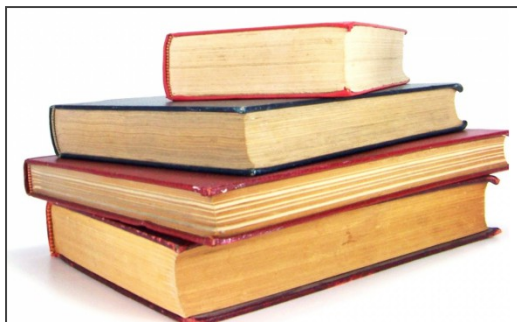
Open Access Database

Molecular Structures

Data Collectors

> **45,000 compounds**

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

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