



because good research needs good data

# Increasing research impact

## The national data registry

Alex Ball

DCC/UKOLN Informatics, University of Bath

11 March 2014



Except where otherwise stated, this work is licensed under the Creative Commons Attribution 4.0 International licence:  
<http://creativecommons.org/licenses/by/4.0/>

Supported by



# UK Research Data (Metadata) Registry Pilot Project

## Project Team

- ▶ **Kevin Ashley**, DCC (Edinburgh)
- ▶ **Alex Ball**, DCC (Bath)
- ▶ **Patrick McCann**, DCC (Glasgow)
- ▶ **Laura Molloy**, DCC (Glasgow)
- ▶ **Veerle Van den Eynden**, UKDA

Funded by **Jisc**



# Outline

Motivation

Project overview

Architecture

Collaborators

Metadata

Evaluation

Future



# UK data landscape



# Motivation

- ▶ Not just specialist data centres any more...
  - ▶ Institutional data repositories
  - ▶ Generalist repositories related to journals
- ▶ Interdisciplinary and multidisciplinary research requires data drawn from diverse sources.
- ▶ Data as a first class research output
  - ▶ Funder impact
  - ▶ Research Excellence Framework



# Brainstorming the national data registry





Browse by Subject Area



Browse by Map Coverage



Advanced Search

## What's in Research Data Australia



### Collections (92633)

Research datasets or collections of research materials.



### Parties (25467)

Researchers or research organisations that create or maintain research datasets or collections.



### Activities (40674)

Projects or programs that create research datasets or collections.



### Services (184)

Services that support the creation or use of research datasets or collections.

## Spotlight on research data

### N.C.W. Beadle Herbarium

The N.C.W. Beadle Herbarium (NE) at University of New England contains around 90,000 pressed, dried, incorporated and databased plant specimens. The collection includes more than 150 TYPE specimens that anchor scientific names as cited in the original publication of those names. This rich resource contains many collections that are of great interest to local and international researchers. The specimen sheet collection of the N.C.W. Beadle Herbarium is databased and available to registered users for online data entry and data query.



[Explore the N.C.W. Beadle Herbarium Collection through Research Data Australia >>>](#)

# Research Data UK?

## Attractions of the Research Data Australia software:

- ▶ Familiar to project team
- ▶ Proven technology
- ▶ Plays nicely with search engines
- ▶ Displays sample citations and access/rights information up front

## Challenges of using the software in the UK:

- ▶ Not used before outside Australia
- ▶ Uses uncommon metadata standard (RIF-CS) internally
- ▶ Original implementation only harvests in RIF-CS
- ▶ No UK data centre can output RIF-CS metadata



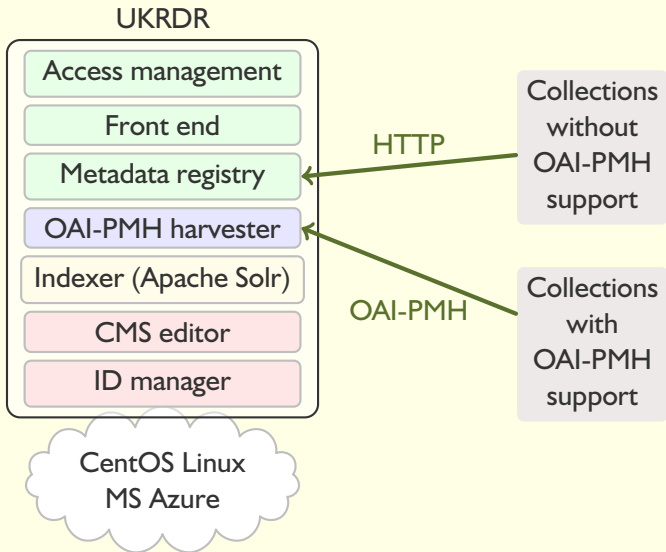


# Project overview

1. Implement a working instance of the ANDS software.
2. Assemble a group of contributors and establish how their metadata will be harvested.
3. Write crosswalks for transforming contributed metadata into RIF-CS.
4. Harvest metadata from contributors.
5. Reports on
  - ▶ using the Research Data Australia software;
  - ▶ how harvesting from data centres went;
  - ▶ how harvesting from university repositories went;
  - ▶ the value of continuing to develop the registry.



# Architecture



# Collaborators

## Data centres:

- ▶ UK Data Archive
- ▶ NERC Data Catalogue Service
  - ▶ BADC
  - ▶ BODC
  - ▶ EIDC
  - ▶ NEODC
  - ▶ NGDC
  - ▶ PDC
  - ▶ UKSSDC
  - ▶ ADS

## Universities:

- ▶ Edinburgh
- ▶ Glasgow
- ▶ Hull
- ▶ Lincoln
- ▶ Leeds
- ▶ Oxford
- ▶ Oxford Brookes
- ▶ St Andrews
- ▶ Southampton



# Metadata crosswalks

## DDI Codebook 2.5

- ▶ UK Data Archive

## DataCite 3

- ▶ Edinburgh (TBC)
- ▶ Oxford (TBC)
- ▶ Hull (TBC)

## OAI-PMH Dublin Core

- ▶ Oxford Brookes (TBC)

## UK Gemini 2.2

- ▶ NERC Data Catalogue Service

## EPrints 3

- ▶ Glasgow
- ▶ Leeds
- ▶ Lincoln (TBC)
- ▶ Southampton

# DDI Codebook 2.5 Crosswalk

RIF-CS 1.5 element	Source (using DDI 2.5 record)
collection[dateAccessioned]	codeBook > stdyDscr > citation > distStmt > depDate[date]
identifier[type=doi]	codeBook > stdyDscr > citation > titlStmt > IDNo[agency=atacite]
identifier[type=local]	codeBook > stdyDscr > citation > titlStmt > IDNo [[agency]:content]
name[type=primary] > namePart	codeBook > stdyDscr > citation > titlStmt > titl
name[type=alternative] > namePart	codeBook > stdyDscr > citation > titlStmt > altTitl
dates[type=dc.available, dc.issued] > date[type=dateFrom]	codeBook > stdyDscr > citation > distStmt > distDate
dates[type=dc.dateSubmitted] > date[type=dateFrom]	codeBook > stdyDscr > citation > distStmt > depDate[date]
location > address > electronic[type=url] > value	codeBook > stdyDscr > citation > holdings[URI]
subject[type=hasset]	codeBook > stdyDscr > stdyInfo > subject > keyword[vocab=S]
subject[termIdentifier]	codeBook > stdyDscr > stdyInfo > subject > keyword[vocab=S; vocabURI]
subject[type=ukdasc]	codeBook > stdyDscr > stdyInfo > subject > topClas
description[type=full]	codeBook > stdyDscr > stdyInfo > abstract
coverage > temporal > date[type=dateFrom]	codeBook > stdyDscr > stdyInfo > sumDscr > colDate[event=start, single; date], timePrd[event=start, single; date]
coverage > temporal > date[type=dateTo]	codeBook > stdyDscr > stdyInfo > sumDscr > colDate[event=end; date], timePrd[event=end; date]
coverage > spatial[type=text]	codeBook > stdyDscr > stdyInfo > sumDscr > geogCover, geogUnit, nation; codeBook > stdyDscr > stdyInfo > subject > keyword[vocab=G]
relatedInfo[type=metadata] > identifier[type=uri]	'http://esds.ac.uk/DDI25/' + UKDA ID + '.xml'

```
<?php
class Ddl2p5toR1fcs extends Crosswalk {
    private $oalpmh = null;
    private $r1fcs = null;
    private $dlfProviders = array(
        "http://oal.ukdataservice.ac.uk/oal/provider" => "UK Data Archive",
    );

    function __construct(){
        require_once(REGISTERED_APP_PATH . "core/crosswalks/_crosswalk_helper.php");
        $this->r1fcs = stplExml_load_string(CrosswalkHelper::RIFCS_WRAPPER);
    }

    public function Identify(){
        return "DDI v2.5 to RIF-CS (Experimental)";
    }

    public function metadataFormat(){
        return "ddl_2_5";
    }

    public function payloadToRIFCS($payload){
        $this->load_payload($payload);
        foreach ($this->oalpmh->listRecords->children() as $record){
            if ($record->getName() != "record") {
                continue;
            }
            $req_obj = $this->r1fcs->addChild("registryObject");
            if (array_key_exists('string') $this->oalpmh->request, $this->ddlProviders()) {
                $req_obj->setAttribute("group", $this->ddlProviders['string'] $this->oalpmh->request);
            }
            $key = $req_obj->addChild("key", $record->header->identifier);
            $originatingSource = $req_obj->addChild("originatingSource", $this->oalpmh->request);
            $coll = $req_obj->addChild("collection");
            $coll->setAttribute("type", "dataset");
            $coll->setAttribute("dateModified", date(DATE_MBC));
            $citation = $coll->addChild("citationInfo");
            $citation_metadata = $citation->addChild("citationMetadata");
            $coverage = $coll->addChild("coverage");
            $rights = $coll->addChild("rights");
            foreach ($record->metadata->codeBook->stdyDscr->children() as $node){
                foreach ($node->children() as $subnode) {
                    $func = "process_".$subnode->getName();
                    if (is_callable(array($this, $func))){
                        call_user_func(
                            array($this, $func),
                            $subnode,
                            array(
                                "registryObject" => $req_obj,
                                "key" => $key,
                                "collection" => $coll,
                                "citation_metadata" => $citation_metadata,
                                "coverage" => $coverage,
                                "rights" => $rights
                            )
                        );
                    }
                }
            }
        }
        return $this->r1fcs->asXML();
    }
}
```

# From UKDA to UKRDR

UK Data Service



## Catalogue

UK Data Service data catalogue record for:

### Attitudes of Students at the London School of Economics, February 1980

[Documentation](#) | [Related Studies](#)

[Download Order](#) | [Get Full DDI XML](#)

#### TITLE DETAILS

**SN:** 1354  
**Title:** Attitudes of Students at the London School of Economics, February 1980  
**Persistent Identifier:** 10.5255/UKDA-SN-1354-1  
**Series:** Attitudes of Students at the London School of Economics, 1980-  
**Depositor:** Husbands, C., London School of Economics and Political Science, Department of Sociology  
**Principal Investigator(s):** Husbands, C., London School of Economics and Political Science, Department of Sociology

#### SUBJECT CATEGORIES

Higher and further

#### ABSTRACT

To conduct a course exercise that collects questionnaire-based information each year from a sample of students at the London School of Economics. The studies focus on background characteristics relevant to a student population, on attitudes to selected political and social issues, and on participation in various activities at LSE. Questions vary somewhat from year to year.

#### COVERAGE, UNIVERSE, METHODOLOGY

**Dates of fieldwork:** 8 February 1980 - 22 February 1980  
**Country:** England  
**Geography:** London  
**Observation units:** Individuals  
Groups  
**Universe:** Subnational  
Students  
A sample of registered part-time and full-time students at London School of Economics and Political Science each year between 1980-1992  
**Time dimensions:** Repeated cross-sectional study  
surveys conducted annually  
**Sampling procedure:** Quota sample  
based on sex, undergraduate/graduate status, domestic/coversness status, and department  
**Number of units:** 288 (single) 280 (coloured)  
**Method of data collection:** Face-to-face interview  
**Weighting:** No information recorded

[Home](#) / [UK Data Archive](#) / [Collection](#)

## Attitudes of Students at the London School of Economics, February 1980

To conduct a course exercise that collects questionnaire-based information each year from a sample of students at the London School of Economics. The studies focus on background characteristics relevant to a student population, on attitudes to selected political and social issues, and on participation in various activities at LSE. Questions vary somewhat from year to year.

#### How to Cite this Collection

##### Citation (Metadata):

Husbands, C. ( 1 Jul 1, Jul 1, Jul 1, Jul 1 ): Attitudes of Students at the London School of Economics, February 1980. UK Data Service. DOI: 10.5255/UKDA-SN-1354-1. <http://dx.doi.org/10.5255/UKDA-SN-1354-1> <http://uk-ids.org/10.5255/UKDA-SN-1354-1>

#### Identifiers

**Local:** an1354  
**DOI:** 10.5255/UKDA-SN-1354-1

#### Additional Metadata

**URL:** <http://eeds.ac.uk/ID25/1354.xml>

#### Spatial Coverage:

**text:** GREATER LONDON  
**text:** England  
**text:** London

#### Temporal Coverage:

From 1980-02-08 to 1980-02-22

#### Subjects

Keywords

ABORTION (INDUCED) ALCOHOL CONSUMPTION ATTITUDES  
EDUCATIONAL FEES EDUCATIONAL FINANCE EDUCATIONAL GRANTS  
FAMILY INFLUENCE FOREIGN STUDENTS GENDER NARCOTIC DRUGS  
OCCUPATIONS PARENTS PART-TIME COURSES  
POLITICAL PARTICIPATION PORNOGRAPHY SEXUAL BEHAVIOUR  
SMOKING SOCIAL ACTIVITIES (LEISURE) SOCIAL CLASS  
SOCIAL PROTEST STUDENT HOUSING STUDENT LEISURE  
STUDENT PARTICIPATION STUDENTS UNIVERSITY COURSES  
Higher and further

#### Access

<http://dx.doi.org/10.5255/UKDA-...>

#### Access rights

The depositor has specified that registration is required and standard conditions of use apply. The depositor may be informed about usage. See terms and conditions of access for further information.

#### Connections

##### People

C. Husbands <sup>PM</sup>

#### Suggested Links

##### Internal Records

288 records with matching subjects

##### External Records

1 records from DataCite

# Evaluation questions

- ▶ Does the software work as intended?
- ▶ Do the harvested records look useful and accurate?
- ▶ Is the system straightforward to use?
- ▶ What might be improved?
- ▶ What additional functions would be desirable?



# Future work

## Formal evaluation

- ▶ ROAMEF = Rationale, Objectives, Appraisal, Monitoring, Evaluation, Feedback

## Questions to consider

- ▶ Would another platform suit us better?
- ▶ Would another internal metadata scheme suit us better than RIF-CS?
- ▶ What use cases should the registry target?
- ▶ How can we add value to the registry's records?
- ▶ Could the registry add value to other systems?







because good research needs good data

Thank you for your attention

DCC Website: <http://www.dcc.ac.uk/>

Alex Ball: <http://alexball.me.uk/>

UKRDR Pilot Project: [http://www.dcc.ac.uk/  
projects/research-data-registry-pilot](http://www.dcc.ac.uk/projects/research-data-registry-pilot)

