Incentivising Uptake of Metadata in the Social and Behavioural Sciences

Louise Corti
Collections Development and Producer Support

EDDI 14
London
December 2014
Through our evangelising we are all incentivising
Britpop, BritArt… Britmetadata
Definition of incentivise (OED)

Motivate or encourage (someone) to do something; provide with an incentive:

Origin

Late Middle English: from Latin incentivum 'something that sets the tune or incites', from incantare‘ to chant or charm'.
My conceptual definitions of Incentivise

- Evangelise
- Demonstrate
- Persuade
- Cajole
- Coerce
- Beg
- Terrify
Operationalising concepts (tactics)

- Evangelise – it really is the best thing since…. You need it in your life
- Demonstrate – Look at this …you’ll want one too!
- Persuasion/cajole – it’ll make you look good
- Encouragement – go on, be an early adopter!
- Coerce – you should be …everyone else is. You’ll look bad if you don’t…
- Beg – Please, it will help us enormously
Source: Boardway.com
THE BOOK OF DDI
Who needs incentivising?

Survey owners and producers

Researchers sharing data

Publishers of data

One’s own organisation
Survey producers

- UK Data Service (very) short brochure
- Closer portal
- Making survey metadata reusable across the lifecycle
- LifeStudy plans
Short brochure for survey products

• Worked closely with data owners and producers

• Existing information too complex

• What is really expected!

• Transferrable information

• Not a bible
Sticks?

- Specify data documentation requirements in the commissioning tender for fieldwork

**Example**
The Centre for Longitudinal Studies (CLS) currently commissions the national cohort studies: the Millennium Cohort Study; the National Child Development Study; the 1970 British Cohort Study and the Longitudinal Study of Young People in England. CLS has started to expect as a deliverable the Computer Assisted Interviewing (CAI) implementation as a Data Documentation Initiative (DDI) compliant XML file, and a file that maps the CAI question to the data variables.

- Mapping between questions and data outputs
- Improved readable questionnaire for end users
CLOSER project

- ESRC/MRC jointly funded discovery service
- 9 long-running birth cohorts & longitudinal studies
- Drivers for project
  - Harmonisation (biomedical, socio-economic)
  - Capacity building
  - Data Linkage
  - Impact
  - Discovery
- Encourage use of existing data resources
Impact and discovery

- DDI 3.2
- Colectica Repository, SDK and Portal
- CADDIES questionnaire entry tool (CLS)
- Manual entry & controlled workflow
CLOSER Incentives

- Jane Elliot key incentive to getting key studies on board
- And some £££
- CLS did all the enhancement work

- Data managers
  - happy to be part of peer group
  - rewarding to go back and look at data
  - liked a shared controlled vocabulary
  - received Colectica training and local installation
  - variable to questionnaire mappings useful
  - liked visibility of their study in the CLOSER platform
Incentivising established survey PIs

- Challenge – to get established/older PIs to recognise the benefits of enhancing metadata
- Midlife study in the US (MIDUS) quite unique!
- Help them ‘See the light’
- For this we need to show something very cool
Variable Details

Concept
Perceived satisfaction

qstnLit
At present, how satisfied are you with your LIFE? Would you say A LOT, SOMEWHAT, A LITTLE, or NOT AT ALL?

preQTxt
And now a few questions about you.

Appears Within

- Studies (1)
- Series (1)
- ( ) Variable Groups (3)
- ( ) Data Files (1)

Ratings

Rating:
Be the first to rate this item.

My Rating
Survey production lifecycle

- Beset with manual processes
- Legacy systems
- Reluctancy to change or adapt systems
- Hard to embrace new ways – disruptive
Typical process

- Manual questionnaire entry (doc/excel/database/XML)
- Export to word
- Deliver to survey agency
- Transfer to SPSS Data Collection
- Export SPSS and PDF/word questionnaire
Metadata publishing

- Data documentation systems and question banks
- Data exploration systems
- Hard to match up Q and Variable information
- So much manual work
- Must do better….
Published outputs – online access


Variable ahlstat: Health over last 12 months - Wave 1

**LITERAL QUESTION**
Please think back over the last 12 months about how your health has been. Compared to people of your own age, would you say that your health has on the whole been excellent, good, fair, poor or very poor?

<table>
<thead>
<tr>
<th>Values</th>
<th>Categories</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>Don't know</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>Excellent</td>
<td>2835</td>
</tr>
<tr>
<td>2</td>
<td>Good</td>
<td>4473</td>
</tr>
<tr>
<td>3</td>
<td>Fair</td>
<td>1801</td>
</tr>
<tr>
<td>4</td>
<td>Poor</td>
<td>601</td>
</tr>
<tr>
<td>5</td>
<td>Very poor</td>
<td>195</td>
</tr>
<tr>
<td>-10</td>
<td>Respondent absent this wave</td>
<td>0</td>
</tr>
<tr>
<td>-9</td>
<td>Missing or wild</td>
<td>1</td>
</tr>
<tr>
<td>-8</td>
<td>Inapplicable</td>
<td>0</td>
</tr>
<tr>
<td>-2</td>
<td>Refused</td>
<td>0</td>
</tr>
</tbody>
</table>

**SUMMARY STATISTICS**
Valid cases: 9911
Missing cases: 1
This variable is numeric
Published outputs – question bank
High Level GSIM

- **Questionnaire design**
  - Questionnaire Specification / implementation validation
  - Output validation
  - Questionnaire capture

- **Data output**
  - Data editing / cleaning
  - Output definition

- **Data ingest**
  - Data validation
  - Business rule validation
  - Security implementation

- **Data processing**
  - Business rule validation
  - Harmonisation

- **Data output**
  - Question / data mapping
  - Questionnaire Documentation
  - Data documentation

- **Data Archiving**
  - Questionnaire Documentation
High Level Statistical Model - GSIM

- **Questionnaire design**
  - Questionnaire Specification / implementation validation
  - Output validation
  - Questionnaire capture

- **Data output**
  - Data editing / cleaning
  - Output definition

- **Data ingest**
  - Data validation
  - Business rule validation
  - Security implementation

- **Data processing**
  - Business rule validation
  - Harmonisation

- **Data output**
  - Question / data mapping
  - Questionnaire Documentation
  - Data documentation

- **Data Archiving**
  - Questionnaire Documentation
  - Data documentation
Gentle XML

- ISER – bespoke Questionnaire Specification Language
- Blaise – import by Michigan Questionnaire Documentation System (MQDS) DDI 3
- CLS using metadata (DDI 2.5)
- CLOSER using DDI 3.2 profile
- UK Data Archive – published metadata (DDI 2.5)
Welcome to our new site. We have changed our name to the Gateway to Global Aging Data and moved our site to g2aging.org. If you previously had an account with us, you will need to re-register as a user. Please enjoy all the features of our new site.

About the Gateway
The Gateway to Global Aging Data is a platform for population survey data on aging around the world. This site offers a digital library of survey questions, a search engine for finding comparable questions across surveys, and identically defined variables for cross-country analysis.

Learn More

What’s New

- Slides from 2014 Partner’s Harmonization Workshop
- Version C of the Harmonized ELSA available
- Introduction to the Gateway to Global Aging Data Webinar
- Check out our booth at this year’s GSA Conference
HEHELF

Location: ELSA » ELSA 2002 » HE. Health
Description: Health status
Item type: Question
Question text: Would you say your health is...
Answer type: Enumerated
Answer choices: 1. Excellent,
2. Very good,
3. Good,
4. Fair,
5. Or, poor?

Harmonized measure: This item is used in a Harmonized survey.
Associated variables: hehelf
Flowchart: locate in flowchart

Topics: health in general
Concurrent items: HEHELF (ELSA 2008), HEHELF (ELSA 2006), HEHELF (ELSA 2004)
The Health module concerns the respondent's state of health, functional limitations, and health behaviors. The main subsections are designed to capture information on individual health status, including self-reported general health, longstanding illness and limiting longstanding illness, eyesight, hearing, walking, awareness of disease, chronic diseases and psychiatric problems ever diagnosed by a doctor, falls, fractures sustained as a result of falls, joint replacement, dizziness, respiratory symptoms, urinary incontinence, disability and functioning, receipt of formal and informal care in connection with IADLs, health behaviors including smoking, drinking, fruit and vegetable consumption, and self-report of level of physical activity.

**Health status**
Would you say your health is...

**Whether has longstanding illness**
Do you have any longstanding illness, disability or infirmity? By longstanding illness I mean anything that has troubled you over a period of time, or that is likely to affect you over a period of time?
Survey Metadata: Barriers & Opportunities
Workshop: 26 June 2014

University Departments and Centres

Market research organisations

Software suppliers and vendors
Meeting outcomes

• Great turn out and knowledge exchange!

• Even… some excitement

• Turn around of principles – quickly

• Tweeting getting good responses

• Planning meetings with UK-IBM and smaller vendors

• Please be an advocate!
Survey Metadata Reusability and Exchange: A Call to Action for Questionnaire Documentation

Last Updated: Wed, 2014-11-05 15:06 — ddiadmin

View Endorsements  Add Endorsement

Produced by the “Survey Metadata: Barriers and Opportunities” Meeting
June 26, 2014, London

Issued 31 October, 2014

This collaborative statement calls upon the survey design, production and archiving communities to take leadership in facilitating survey metadata exchange through adoption of shared metadata standards for questionnaire and data description.

Background

Statement of Shared Principles for Survey Metadata Reusability and Exchange

Currently there are some obstacles in creating good structured documentation for data that complies with international gold standards for enabling use and longer-term re-use.

The journey from questionnaire design using commercial software to Computer Assisted Interviewing (CAI) delivery in the field, and to delivery of data to owners and archives is currently inefficient and error-prone. The metadata journey from owner to producer back to owner and to archive continues to use
New Life Study

- UK-wide birth cohort study
- Data, environmental and biological samples
- Study design
  - 60,000+ pregnant mothers and their partners recruited through selected maternity units
  - birth component - nationally representative sample of 20,000 babies recruited through the birth register
- Biological samples
  - from the mothers and their partners and babies
  - at delivery – ‘cutting-edge’ samples - relevant to assessment of infection, immunity and the microbiome
Life Study survey design

- Much thought upfront about reusability of questionnaire metadata
- New cohort study, so opportunities
- Bespoke data collection system using XML schema
- Incentivised by previous (local) poorer metadata practice
Researchers ...messy long-tail data

- 20 years of dreaded Data Policy
- Self-deposit repository
- Jisc Managing Research Data Programme local pilots and activities
- Review of data and incentives
ESRC research data policy

Research data should be openly available to the maximum extent possible through long-term preservation and high quality data management. (ESRC Research Data Policy, 2010)

• ESRC grant applicants planning to create data during their research include a data management plan

• ESRC award holders offer their research data to UK Data Service within three months of the end of their grant

Researchers who collect the data initially should be aware that ESRC expects that others will also use it, so consent should be obtained on this basis and the original researcher must take into account the long-term use and preservation of data. (ESRC Framework for Research Ethics, 2012)
UK Data Service ReShare

http://reshare.ukdataservice.ac.uk/

• A platform for sharing research data
• Data produced by UK social science researchers, and data of interest to social science researchers
Easy to publish and upload data

Edit collection: Data Collection #851515

To deposit a data collection, you must accept the ReShare Terms and Conditions.

- I confirm that I am the owner of the copyright and associated intellectual property rights in the whole Data Collection or am otherwise lawfully entitled to grant this licence on behalf of each and every owner;
- I grant a non-exclusive, royalty-free licence to the UK Data Archive (a department of the University of Essex and not a separate legal entity) of Wivenhoe Park, Colchester, CO4 3SQ (the "University") to hold, make copies of, and disseminate copies of the Data collection, in accordance with the access conditions I will specify when uploading data files: open data accessible to users without registration, or safeguarded data accessible to users registered with the data services provided by the UK Data Archive.

I agree to the ReShare data deposit terms and conditions

Save for later  Cancel  Next >
Collection period: 1 October 2008 to 30 September 2012
Country: United Kingdom
Data collection method: Collection of information from official documentation across several countries
Observation unit: Organisations, Text units
Kind of data: Alpha-numeric
Type of data: Qualitative and mixed methods data
Resource language: English

Access and Administration

Data sourcing, processing and preparation: Information from various sources, including the International Bureau of Fiscal Documentation
Copyright holders:
<table>
<thead>
<tr>
<th>Name</th>
<th>Email</th>
<th>Affiliation</th>
<th>ORCID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Devereux, Michael</td>
<td>Unspecified</td>
<td>University of Oxford</td>
<td>Unspecified</td>
</tr>
</tbody>
</table>
Contact:
<table>
<thead>
<tr>
<th>Name</th>
<th>Email</th>
<th>Affiliation</th>
<th>ORCID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Devereux, Michael</td>
<td><a href="mailto:michael.devereux@sbs.ox.ac.uk">michael.devereux@sbs.ox.ac.uk</a></td>
<td>University of Oxford</td>
<td>Unspecified</td>
</tr>
</tbody>
</table>
Notes on access: None
Publisher: Economic and Social Research Council
Last modified: 28 Apr 2014 22:42

Available files:
- CBT_Tax_database.xlsx
- CBT_Tax_database_description.pdf
<table>
<thead>
<tr>
<th>Variables:</th>
<th>Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Format</td>
<td>Description</td>
</tr>
<tr>
<td>country</td>
<td>3 digit string</td>
<td>3 digit ISO Code</td>
</tr>
<tr>
<td>year</td>
<td>4 digit number</td>
<td>Calendar year (NOT fiscal year)</td>
</tr>
<tr>
<td>top_corp_tax_rate</td>
<td>Numeric, not in %</td>
<td>Top statutory corporate tax rate at central government level – excluding eventual surcharges, on distributed profits</td>
</tr>
<tr>
<td>surcharge</td>
<td>Numeric, not in %</td>
<td>Surcharges on corporate tax</td>
</tr>
<tr>
<td>local_top_corp_tax_rate</td>
<td>Numeric, not in %</td>
<td>Top statutory corporate tax rate at lower government levels</td>
</tr>
<tr>
<td>local_tax_deductible</td>
<td>Numeric, not in %</td>
<td>This variable takes value 0 if no local tax is applicable or local corporate tax payments are NOT deductible from central government corporation tax, 1 if local taxes are deductible from central taxes, e.g. 0.5 if 50 percent are deductible</td>
</tr>
<tr>
<td>local_is_average</td>
<td>Dummy</td>
<td>0 if no local tax or if the local tax of a particular place is used. 1 if the local corporate tax rate is an average of all provinces</td>
</tr>
<tr>
<td>local_is_typical</td>
<td>Dummy</td>
<td>0 if no local tax or if the local tax an average of all provinces. 1 if the local corporate tax rate is for a typical place (mostly economic capital)</td>
</tr>
<tr>
<td>total_corp_tax_rate</td>
<td>Numeric, not in %</td>
<td>Sum of federal tax rate, local tax rate taking into account surcharge and deductibility of local taxes</td>
</tr>
</tbody>
</table>
Self-upload guidance....

- Lots of it…guides, webinars, hand-holding

- Still some poor metadata
  - file naming; data inventories; variable names

- Review criteria are explicit

- More comms work to do!

- Incentives coming through ‘star’ quality rating
Guidance on data review

ReShare Data Review

Documentation

The UK Data Service reviews each data collection submitted to the ReShare repository for disclosure risk, copyright breaches, validity of file formats and level of documentation before publishing the collection. The quality or scientific validity of research data is not evaluated.

All numerical data files and at least a 10% random sample of textual data files are reviewed.

Non-English language data files are checked if a staff member of the UK Data Service is familiar with the language.

Any concerns, changes needing to be made to data files, or requests for additional documentation are communicated to the data depositor for actioning.

Data review comments and actions taken are noted in the data collection metadata record, as provenance information for the data collection.

Review procedures

1. Generic project-level checks:
Qualitative research metadata

- Years of evangelism!
- Qualitative metadata standard QuDEx
- Successful implementation in QualiBank
- Working with qualitative researchers
- DDI working group model
Interview with Mr. Keble

DETAILS

Collection ID (SN): 2000
Title: Family Life and Work Experience Before 1918, 1870-1973
Principal Investigator: Thompson, P., University of Essex. Department of Sociology
Lummis, T., University of Essex. Department of Sociology
Sex: Male
Age group: 65-74
Socio-economic status: Semi-routine
Region: East of England
Licence: Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Licence.

Additional metadata

Idno: 2000int001
No. pages: 55
Placename: Colchester, Essex
Date of interview: 1950
R marital status: Married
R occupation: Postman

INTERVIEW TRANSCRIPT
**Details**

**Collection ID (SN):** 2000  
**Title:** Family Life and Work Experience Before 1918, 1870-1973  
**Principal investigator:** Thompson, P., University of Essex. Department of Sociology  
Lummi, T., University of Essex. Department of Sociology  
**Historical note:** Movement of rural populations to urban centres led to the decline of many small-scale village tradesmen. However larger villages could still support business such as wheelwrights and blacksmiths that were central to the local economy.  
**Image description:** Three generations of a family pose for a photograph in front of the family home and business c.1904. Behind them are various works in progress for these wheelwrights including an urban district council wheelbarrow.  
** Licence:** Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Licence.

**Image**

![Image of a family posing in front of a house and business, with various works in progress for wheelwrights including an urban district council wheelbarrow.](image-url)
Incentives

• QualiBank system very simple and appealing
• Usability easily understood
• Value of metadata appreciated
• Uptake of rich object-level metadata
Data publishers

- Data Centres
  - Inspiring catalogues
  - Online browsing systems

- Institutional Repositories

- Journals
  - Transparency agenda and data review

- Worth investing in high quality
Discover at UK Data Service

Discover

Search and browse our data collections, support guides, case studies, and related publications.

Search our data catalogue and related resources

Case study  Data collection  Series record  ESRC output  Support guide

SEARCH RESULTS

Displaying 1-10 of 6361 results

1. SN 2000087 National Awareness and Early Diagnosis Initiative: Cancer Awareness Measures
   Description:
   The purpose of the Cancer Awareness Measure (NAEDI-CAM) is to facilitate accurate and sensitive surveillance of public awareness of cancer over time, identify information needs,...

2. SN 2000086 English House Condition Survey
   Description:
   The English House Condition Survey (EHCS) was a national survey of housing in England, commissioned by the Department for Communities and Local Government (DCLG). The survey was...

3. SN 2000085 National Cancer Patient Experience Survey
   Description:
## Online instant data browsing

<table>
<thead>
<tr>
<th>Nesstar</th>
<th>social surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>UKDS.stat</td>
<td>aggregate global indicators</td>
</tr>
<tr>
<td>InFUSE</td>
<td>aggregate census data</td>
</tr>
<tr>
<td>QualiBank</td>
<td>qualitative data</td>
</tr>
</tbody>
</table>
**Variable PartyIDN: R's political party identification :Q251**

### Literal Question

IF 'yes' AT [SupParty] OR AT [ClosePty]: Which one? IF 'no' OR DON'T KNOW AT [ClosePty]: If there were a general election tomorrow, which political party do you think you would be most likely to support?

<table>
<thead>
<tr>
<th>Values</th>
<th>Categories</th>
<th>N</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Conservative</td>
<td>818</td>
<td>25.2%</td>
</tr>
<tr>
<td>2</td>
<td>Labour</td>
<td>1052</td>
<td>32.4%</td>
</tr>
<tr>
<td>3</td>
<td>Liberal Democrat</td>
<td>194</td>
<td>6.0%</td>
</tr>
<tr>
<td>4</td>
<td>Scottish National Party</td>
<td>55</td>
<td>1.7%</td>
</tr>
<tr>
<td>5</td>
<td>Plaid Cymru</td>
<td>15</td>
<td>0.5%</td>
</tr>
<tr>
<td>6</td>
<td>Green Party</td>
<td>34</td>
<td>1.0%</td>
</tr>
<tr>
<td>7</td>
<td>UK Independence Party (UKIP)/Veritas</td>
<td>181</td>
<td>5.6%</td>
</tr>
<tr>
<td>8</td>
<td>British National Party (BNP)/National Front</td>
<td>28</td>
<td>0.9%</td>
</tr>
<tr>
<td>9</td>
<td>RESPECT/Scottish Socialist Party (SSP)/Socialist Party</td>
<td>6</td>
<td>0.2%</td>
</tr>
<tr>
<td>10</td>
<td>Other party (WRITE IN)</td>
<td>197</td>
<td>6.1%</td>
</tr>
<tr>
<td>11</td>
<td>Other answer (WRITE IN)</td>
<td>15</td>
<td>0.5%</td>
</tr>
<tr>
<td>12</td>
<td>None</td>
<td>581</td>
<td>17.9%</td>
</tr>
<tr>
<td>20</td>
<td>Refused to say</td>
<td>57</td>
<td>1.8%</td>
</tr>
<tr>
<td>98</td>
<td>Don't know</td>
<td>197</td>
<td>6.1%</td>
</tr>
<tr>
<td>99</td>
<td>Refusal</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>-1</td>
<td>Not applicable</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

### Summary Statistics

Valid cases 3242  
Missing cases 2  
This variable is numeric

### Interviewer Instructions

DO NOT PROMPT

### Universe

IF 'yes' AT [SupParty] OR 'yes', 'no' OR DON'T KNOW AT [ClosePty]
Institutional Repositories

• Huge movement in UK. ££££ Jisc MRD Programme
• Pilots and some now set up and mature
• 2 years + to agree reach software and metadata requirements
• Lots of arguing over minimal requirements
  Politics – Funder or not?
• Lots of standing up and saying DDI
## 1834 Parisian deaf-mute banquet

<table>
<thead>
<tr>
<th>Alternative title</th>
<th>Banquet des sourds-muets parisien 1834</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creator(s)</td>
<td>Michael Gulliver</td>
</tr>
<tr>
<td>Funder</td>
<td>Economic and Social Research Council</td>
</tr>
<tr>
<td>Publication date</td>
<td>15 Aug 2013</td>
</tr>
<tr>
<td>Language</td>
<td>eng</td>
</tr>
<tr>
<td>Publisher</td>
<td>University of Bristol</td>
</tr>
<tr>
<td>DOI</td>
<td>10.5523/bris.1ghwfnnggawvoo1qyih0xomzemp</td>
</tr>
<tr>
<td>Compressed dataset</td>
<td><a href="http://data.bris.ac.uk/datasets/1ghwfnnggawvoo1qyih0xomzemp/1ghwf">http://data.bris.ac.uk/datasets/1ghwfnnggawvoo1qyih0xomzemp/1ghwf</a></td>
</tr>
</tbody>
</table>
Research Data Registry and Discovery

- JISC pilot project to provide a coherent point of access to descriptions of UK research datasets

- Adapted Australian National Data Service (ANDS) software (ORCA) for Research Data Australia

- Metadata profiles used by data centres and institutional repositories mapped to Registry Interchange Format – Collections and Services (RIF-CS)

- Metadata imported via OAI-PMH or other modes
What's in Research Data Australia

Collections (99008)
Research datasets or collections of research materials.

Parties (25701)
Researchers or research organisations that create or maintain research datasets or collections.

Activities (40731)
Projects or programs that create research datasets or collections.

Services (187)

Spotlight on research data

CATAMI:
Collaborative and Annotation Tools for Analysis of Marine Imagery and Video

In recent years, ecologists and biologists have increasingly relied on remote video and image-based methods to examine and monitor marine habitats, particularly those beyond easy diving depth. These data streams are currently being collected by the end user community at sites around Australia using modern image and video based tools including Baited Remote Underwater Video Systems (BRUVS), Autonomous Underwater Vehicle (AUV), Remotely Operated Vehicle (ROV) and Underwater Towed Video (UTV) systems. Transforming raw visual data into quantitative information useful for science and policy decisions requires substantial effort by human experts. The University of Sydney has led a collaborative project to develop common labelling standards, and to transform and collate existing datasets that support marine research across Australia. The result is CATAMI: an online data management and image annotation system for marine ecologists.

Access the CATAMI service through Research Data Australia...
PHP crosswalks

- DDI Codebook 2.5
- UK GEMINI.2
- OAI Dublin Core (oai_dc)
- EPrints (ReCollect) metadata export scheme
- DataCite v3 from DataCite’s OAI-PMH endpoint;
- MODS V3.5 from DSpace-powered OAI-PMH endpoints

- RIF-CS not ideal
- CKAN being explored
Incentivising new registry contributors

- Pilot repository as a showcase
- Recognition by new contributors of metadata requirements
- Poor metadata looks shoddy!
- Training useful
Journals

- Science journals have data policies for data sharing
  - Science, Nature, PLOS ONE
    “PLOS ONE will not consider a study if the conclusions depend solely on the analysis of proprietary data” … “the paper must include an analysis of public data that validates the conclusions so others can reproduce the analysis.”
  - BioMed Central open data statement

- Data underpinning publication accessible via a DOI
  - upon request from author
  - as supplement with publication
  - in public repository
  - in mandated repository (PANGAEA – Elsevier)
Journals

• Social science needs some motivation

• American Political Science Association (APSA) Council - policies guiding data access and research transparency in political science

• Endorsement by journals

• Training in how to prepare and submit supporting data and sufficient metadata
Peer review of data

- Data centres do this - but not for quality of methods
- Journals doing this - replicability agenda
- No one single standard for ‘quality’
- Make metadata quality explicit:
  - Collection description
  - Data description: file and variable names & labels
  - Relationships between tables/files
Incentivising our staff

- Not so many into DDI

- ‘Language’ barrier

- CESSDA expert seminars to share problems

- Engage in data processing and producer relations work
Some tips on incentivising

- Speak a common language
- On DDI, do say it is stable and offer one version only
- Start with the lowest common denominator. Baby steps
- Show value – shiny interfaces and examples!
- Where possible agree to help with the (initial) work
- Integrate into everyday workflows and research tools
CONTACT

UK Data Service
University of Essex
Wivenhoe Park
Colchester
Essex CO4 3SQ

T +44 (0)1206 872145
E corti@essex.ac.uk