



EMPLOYING **DDI** AT THE **UK DATA ARCHIVE** – NOW AND NEXT

**JACK
KNEESHAW**

SERVICE MANAGER
UK DATA ARCHIVE
UNIVERSITY OF ESSEX

MRC DATA MANAGEMENT WORKSHOP
11 AUGUST 2011





THIS PRESENTATION

- Introduction to the UK Data Archive
- Now: How we use DDI
- Next: Leveraging DDI-Lifecycle (DDI 3)
- Case study: Re-using metadata for survey questions
- In summary

INTRODUCTION TO THE UK DATA ARCHIVE

The screenshot displays the UK Data Archive website. At the top right, there are links for 'HELP', 'CONTACT US', and 'SIGN UP'. Below this is a navigation menu with 'HOME', 'ABOUT US', 'CREATE & MANAGE DATA', 'DEPOSIT DATA', 'HOW WE CURATE DATA', 'FIND DATA', and 'NEWS & EVENTS'. The 'ABOUT US' page is active. A video player shows a video titled 'ABOUT US' with a play button and a progress bar. To the right of the video, text reads: 'We acquire, curate and provide access to the UK's largest collection of social and economic data. Find out more about us'. Below the video, there are sections for 'ABOUT THE ARCHIVE', 'PARTNERS & COLLABORATORS', 'OUR PROJECTS', and 'ACROSS THE DECADES', each with a brief description and a 'READ ON' button. On the right side, there is a search bar, a 'DEPOSIT YOUR DATA' button, a 'FIND DATA' button, and a 'WATCH A VIDEO ABOUT US' button. Below these are sections for 'A QUICK GUIDE TO THE ARCHIVE' and 'OUR SERVICES', with a 'SECURE DATA SERVICE' logo. At the bottom right, there are links for 'HAVE A QUESTION - CONTACT US', 'ABOUT US FAQ', and 'JOIN OUR MAILING LIST'. The left sidebar contains a 'UK DATA ARCHIVE' logo and a list of navigation links: 'ABOUT US', 'ABOUT THE ARCHIVE', 'OUR CORE SERVICES', 'OUR PROJECTS', 'SPONSORS', 'PARTNERS & COLLABORATORS', 'REPORTS & PUBLICATIONS', and 'OUR PEOPLE'.

NOW: HOW WE USE DDI

- Archive has been working with DDI since its inception and has produced standardised study descriptions for catalogue records since the 1970s
- At 2011, we prepare and deliver DDI descriptions through our data catalogue and via Nesstar
- Archive on the Technical Implementation Committee and the DDI Controlled Vocabulary subcommittee
- Worked with ODaF to build an SPSS to DDI converter (DExT) proof of concept focusing on SPSS as input data format with data export capabilities to ASCII, SAS, Stata, and SPSS
- More recent work with DDI Alliance on describing qualitative data



CURRENT DDI WORKFLOW AT THE ARCHIVE

- Depositor completes Data Deposit Form, information imported into relational database
- Data processors add metadata (including HASSET terms) via bespoke input forms
- DDI 2.1-compliant XML record created and stored (as a string)
- Publish DDI 2.1 (with additions/customisations) metadata record to our data catalogue (levels 1-2 only as object is study)
- Index XML record for website Lucene catalogue search
- Run SPSS to DDI transformation script to create variable-level metadata for catalogue (not Nesstar)
- DDI 2.1 record imported into Nesstar Publisher and DDI variable-level metadata added
- Publish DDI 1.2 record in Nesstar catalogue

NEXT: LEVERAGING DDI-LIFECYCLE (DDI 3)

- 2011-12: Review and overhaul of legacy metadata used through ingest, preservation and dissemination (IPD) functions
- Migrating to DDI-Lifecycle as native structure so that metadata subsets can be extracted (and converted to appropriate DDI version) on demand
- **Stage 1:**
 - (i) audit of legacy metadata used in IPD (“haves”)
 - (ii) 4 proofs of concept requirements (“wants”)
- **Ongoing stage 1 and 2:** platform/application evaluation
- **Stage 2:**
 - (i) agreed UKDA DDI-Lifecycle profile
 - (ii) agreed UKDA full metadata profile (DDI-Lifecycle +)
- **Stage 3:** database requirements > applications development

ISSUES TO BE ADDRESSED

- DDI-Lifecycle will not provide home for all elements of IPD
- Even if it does, we may still choose to:
 - (i) manage metadata in DDI 3.x but export to other standards (e.g. export a PREMIS record and maintain an AIP)
 - (ii) manage some metadata outside DDI 3.x (e.g. manage preservation metadata in PREMIS)
- Metadata management must maintain compliance with the moving target of advances to standards (e.g. upgrading to future versions of DDI) and also retain backwards compatibility for existing use/users
- Storage: Relational vs. XML vs. hybrid databases
 - XML format aids search, update, retrieval
 - Hold controlled vocabularies (CVs) in XML for ease of integration
- Many, many third party tools to be evaluated!
<http://www.ddialliance.org/resources/tools>

“WANTS”: PROOFS OF CONCEPT

- HASSET
 - Social science thesaurus – in essence a CV
 - Publish as a resource package? Maintainable object with UKDA as agency?
- Geo-referencing/INSPIRE
 - INSPIRE - Infrastructure for Spatial Information in the European Community (European directive 2007/2/EC)
 - enables exchange, sharing, access to and use of environmental spatial data
 - facilitate public access to spatial information across Europe
 - metadata mapping (by Relu Data Support Service):
 - UKDA DDI 2.1 - INSPIRE - DDI 3.1
 - INSPIRE metadata less detailed than DDI – overall good correspondence
- Longitudinal studies
 - Variable version histories across waves?
 - Variable comparisons with international equivalents?
- Question bank

CASE STUDY: RE-USING METADATA FOR SURVEY QUESTIONS

- Survey Question Bank:
<http://www.surveynet.ac.uk/sqb/qb/questions.asp>
- Front end to ESDS Nesstar catalogue:
<http://nesstar.esds.ac.uk/webview/>
- Searches across the text from questions used in a selection of key, recent (and mainly) UK social surveys
- 150,000+ *questions* across 50+ survey series
- But niggling issues remain based on limitations of *variable*-level DDI 1.2 metadata ...

CASP-19 QUALITY OF LIFE MEASURE

Used by:

- British Household Panel Survey
- English Longitudinal Study of Ageing (plus SHARE)
- National Child Development Study
- (others)

1 Here is a list of statements that people have used to describe their lives or how they feel. We would like to know how often, if at all, you think they apply to you.

Tick one box on each line

	Often	Some- times	Not Often	Never	
My age prevents me from doing the things I would like to	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	218
I feel that what happens to me is out of my control	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	218
I feel free to plan for the future	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	220
I feel left out of things	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	221
I can do the things that I want to do	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	222
Family responsibilities prevent me from doing what I want to do	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	222
I feel that I can please myself what I do	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	224
My health stops me from doing things I want to do	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	225
Shortage of money stops me from doing the things I want to do	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	226
I look forward to each day	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	227
I feel that my life has meaning	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	228
I enjoy the things that I do	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	228
I enjoy being in the company of others	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	230
On balance, I look back on my life with a sense of happiness	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	231
I feel full of energy these days	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	232
I choose to do things that I have never done before	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	232
I feel satisfied with the way my life has turned out	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	234
I feel that life is full of opportunities	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	235
I feel that the future looks good for me	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	236

EXAMPLE SEARCH

Useful but:

- How can I be sure these are identical questions?
- How do I know whether the question belongs to a wider battery or not?

QUESTION BANK

Find survey questions containing

Question text*

* = required field
Enter text that you hope to find in the question wording itself (e.g., **how satisfied**) rather than underlying concepts (e.g., **subjective well-being**).

Refine the search

Survey series

Year From: To:

Results 1 to 2 of 2

Here is a list of statements that people have used to describe their lives or how they feel. We would like to know how often, if at all, you think this applies to you **I feel left out of things.**

National Child Development Study: Sweep 8, 2008-2009: First Deposit
[Access full detail](#)

[Self-completion] **I feel left out of things**

English Longitudinal Study of Ageing (ELSA); Wave 1, 2002-2003: Teaching Dataset
[Access full detail](#)

Results 1 to 2 of 2

EXAMPLE QUESTION RECORD

Currently clumsy:

- Has this question been used before by this study?
- Has this question been used elsewhere?

What else would be nice?:

- What other questions share the same metadata? Shared keywords, shared question type/intent, shared response categories etc.

QUESTION BANK

[<< Previous question](#)

[Next question >>](#)

Question text: Here is a list of statements that people have used to describe their lives or how they feel. We would like to know how often, if at all, you think this applies to you? I feel left out of things.

Responses:

- 1 Item not applicable
- 7 Polar Multicode (7)
- 8 Multicode (8)
- 9 Not answered (9)
- 1 Often
- 2 Sometimes
- 3 Not often
- 4 Never

Survey title: National Child Development Study: Sweep 8, 2008-2009: First Deposit

[View response frequencies for this question](#)



[View this question in other years of the same survey series](#)

QUESTION BANK

[Access full survey dataset](#)



[View the same question in other surveys](#)

QUESTION BANK

[View this question in original questionnaire\(s\) Beta](#)

QUESTION BANK

[Search for this question elsewhere:](#)

ICPSR

WHY A DDI-L QUESTION BANK?

- Benefits for data managers/archives/service providers
 - Internal to a study:
 - Allows for separation of question-level metadata from variable-level metadata
 - Ability to re-use and share metadata
 - Minimise redundancy and discrepancy
 - External:
 - Ability to push out/pull in comparative information
- Benefits for researchers
 - Potential for a cross-repository search interface
 - If not a single search interface then a shared search lingua franca
- Benefits for survey commissioners/fieldwork agencies
 - Ability to identify and re-use questions (promoting harmonisation)
 - Greater visibility of their questions/instruments (impact)

HOW A DDI-L QUESTION BANK?

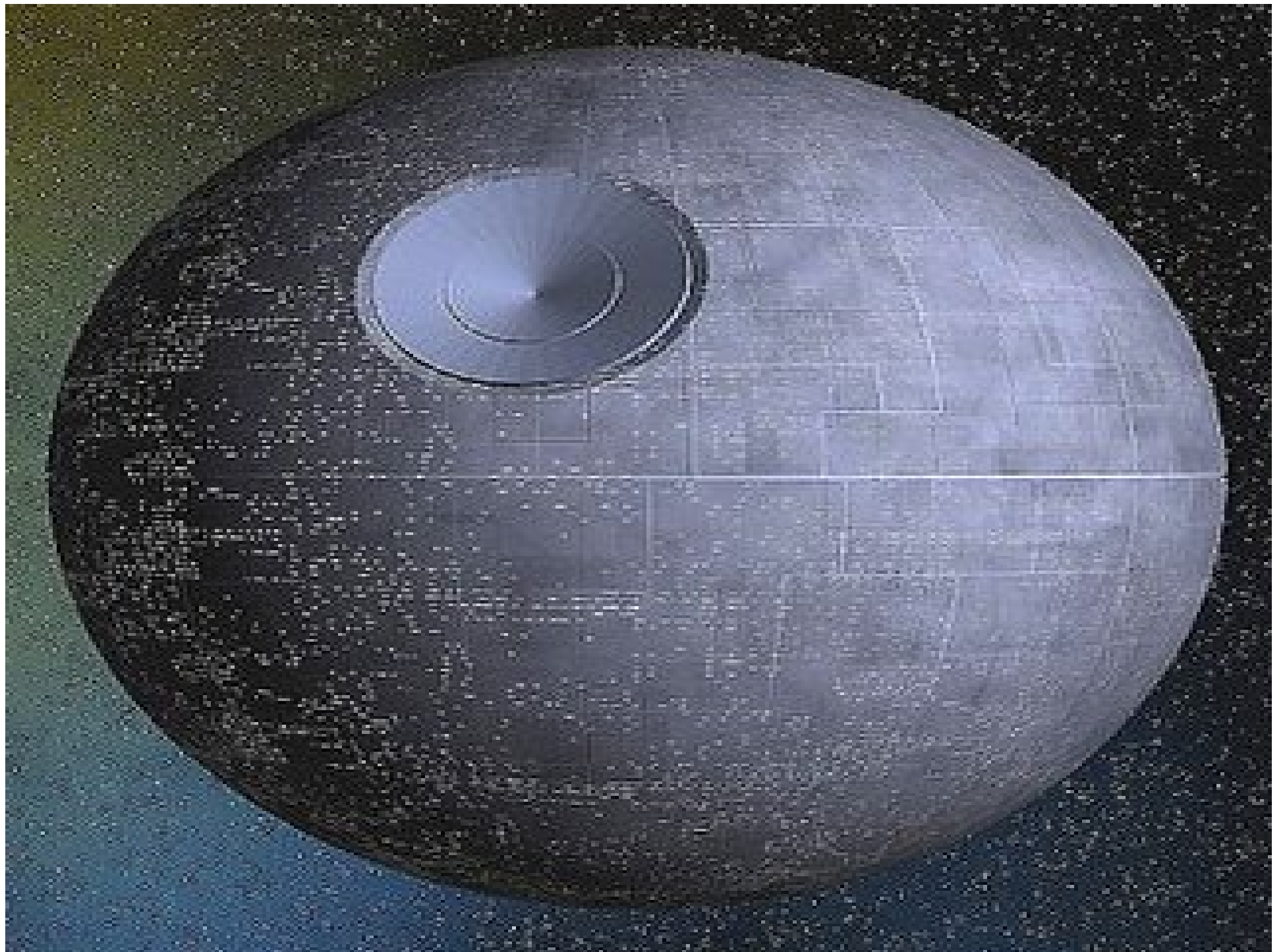
- Opportunity to mark-up:
 - Survey instruments, harmonised question sets and intersections between these <questionScheme>
 - Question version histories <questionItem>
 - Question concept <conceptScheme>
 - Question type/intent/purpose <questionIntent>
 - Question comparator <questionMap>
 - Question context/mode/category/response type <responseDomain> <categoryScheme> etc.
- Do so using DDI schemes and controlled vocabularies (published registries)
 - Promote re-use
 - Explicitly comparable (avoiding clumsy matching, e.g. CASP-19)
 - Maintained by a reputable agency who manages versioning
 - Every scheme has a URN with full versioning:
urn:ddi:<agency>:QuestionScheme.c1.1.0.0:QuestionItem.q1.1.0.0

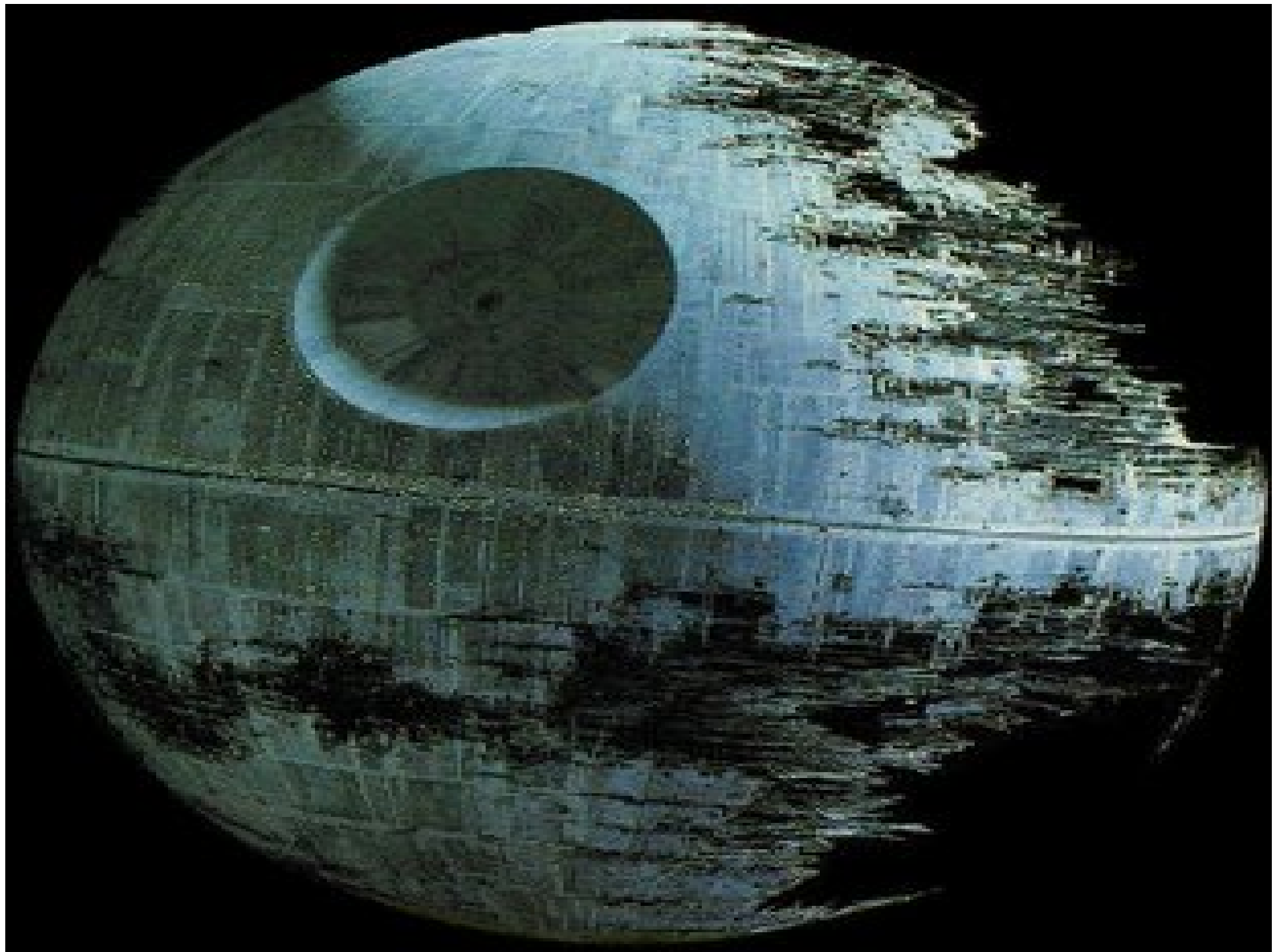
ISSUES?

- DDI-Lifecycle offers much but represents a steep learning curve
 - Helped by use cases/projects at DDI Alliance site
- Lack of published (question) schemes:
<http://registry.ddialliance.org/>
- Much of the work on questions/questionnaire development is best done upstream in the data lifecycle – how do we get the researchers'/data collectors' buy in?
- There are a growing number of tools to make life easier but these need evaluating ...

IN SUMMARY

- Migration to DDI-Lifecycle in early stages but will involve:
 - Substantial workflow changes across all Archive processes, including re-implementation of our input forms
 - Rationalisation and migration of existing metadata to new platform
 - New mindset: Archive is part of the lifecycle not just the publisher of the codebook
- Expected benefits:
 - Potential for greatly improved resource discovery
 - Re-use of schemes/resource packages > minimise redundancy/error
 - Input to new features – e.g., visualisation tools, especially via improved geo metadata







LINKS AND CONTACT

<http://www.esds.ac.uk/findingdata/aboutcatfields.asp>

<http://www.surveynet.ac.uk/sqb/qb/questions.asp>

<http://nesstar.esds.ac.uk/webview/index.jsp>

JACK KNEESHAW
kneejw@essex.ac.uk

UK DATA ARCHIVE
UNIVERSITY OF ESSEX
WIVENHOE PARK
COLCHESTER
ESSEX CO4 3SQ

T +44 (0)1206 872001
E info@data-archive.ac.uk
www.data-archive.ac.uk

UK DATA ARCHIVE