



THE UK DATA ARCHIVE **RESOURCE** **DISCOVERY PROJECT**

**LUCY
BELL**

MANAGEMENT INFORMATION MANAGER
UK DATA ARCHIVE
UNIVERSITY OF ESSEX

ESRC RESOURCE DISCOVERY WORKSHOP, **RIBA**

22 MARCH 2011



**UK • DATA
ARCHIVE**

How do researchers find data in 2011?

- Trends in information-seeking behaviour show that users prefer simple, Google-like interfaces...
- ...but which still return acutely-focused and highly-relevant results.
- The look and feel of the interfaces should be simple but the results must achieve academic rigour.

Metadata

- The power behind the resource discovery systems lies with metadata.
- Three types of metadata:
 - Descriptive metadata (metacontent)
 - Keywords and the contents of the indexed fields – title, author, abstract etc
 - Structural metadata
 - The design and specification of metadata structures, for example, how a book is ordered into chapters or how a data collection is made up of various studies
 - Administrative metadata
 - Technical information, such as file type, date of creation, rights and preservation
- All are important in enabling resource discovery

The metadata conundrum

- For data services to produce simple interfaces - which still return highly-relevant results - metadata are required which are both:
 - Extremely powerful
 - Increasingly invisible
- A challenge exists for services to create resource discovery interfaces which are supported by **high-functioning, yet hidden, metadata.**

Applying the problem to a service

- The UK Data Archive, based at the University of Essex, is the curator of the largest collection of digital data in the social sciences and humanities in the United Kingdom. It holds several thousand datasets relating to society, both historical and contemporary. It has been in existence for over 40 years.
- May – October 2010: a review was carried out of the UK Data Archive's resource discovery tools.

Reviewing the UK Data Archive's tools

- The primary resource discovery tools reviewed:
 - ESDS (www.esds.ac.uk)
 - UK Data Archive (www.data-archive.ac.uk)
 - Census Portal (www.census.ac.uk)
 - Survey Question Bank (www.surveynet.ac.uk)
 - History Data Service (hds.essex.ac.uk)
 - RELU-DSS (relu-data-archive.ac.uk)
 - Controlled vocabularies (HASSET and subject categories)
 - Plus a comparison of resource discovery elsewhere
- The brief:
 - review the Archive's resource discovery tools making recommendations on how they can be improved;
 - undertake a detailed examination of the website;
 - describe each method of resource discovery;
 - document all existing internal procedural documents, standards documents and current user guidance noting gaps;
 - collate evidence from existing reviews about user satisfaction and usability.
- White Paper was submitted to the Archive and ESDS SMTs in November 2010.

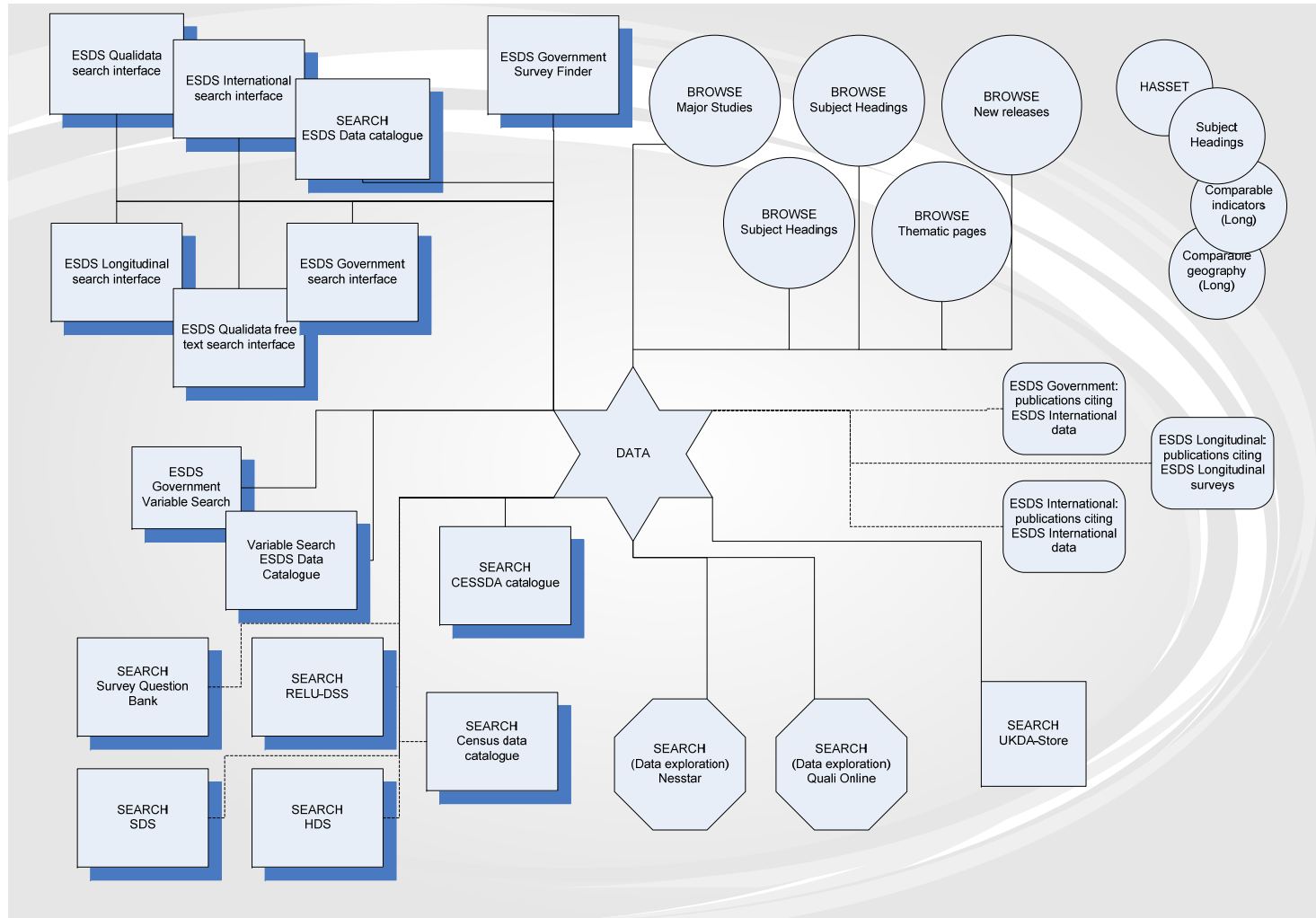
User consultation

- Evidence of user satisfaction was examined:
 - 67% of respondents to a recent survey were satisfied or very satisfied with the ease of navigation
 - 64% were satisfied or very satisfied with the quality of the metadata but 79% of respondents thought that improvements to metadata should be given medium, high or very importance
 - Comments made regarding users having to make 'sudden leaps from one part to another which can make redirection quite difficult'

Reviewing the existing tools

- The top line results:
 - Archive and ESDS users prefer to search, rather than browse
 - ESDS in particular provides many different routes in to the data for its users, and caters for many different types of user (however, users do not always use the different routes appropriately)
 - Dead-ends can exist

The wider Archive context: 21 interfaces



The issue:

- *How to integrate these interfaces and the metadata?*
- The vision:
- **A series of pathways, from a single starting point, allowing the user to move seamlessly from one service/type of data to another.**
- No matter which search is undertaken, the user will be able to move to one of the other areas, and leave with what they want.
- It means linking ESDS data collections to publications, to research outputs, to citations, to question text to web pages and beyond
- Building on work which allows **cross-service searching** and makes **linkages** between data held in different services
 - from question text in SQB to data visualisation in Nesstar
 - from RELU-DSS research outputs to data

SQB: linking questions to data

- SQB tool allows users to search across a selection of key, recent UK social surveys held in the ESDS Nesstar Data Catalogue
 - around 50 survey series
 - more than 200,000 questions (variables)
- (Freetext) search/retrieve questions and link results to:
 - the survey response (frequency counts)
 - the question as it appears in its native questionnaire
 - the full survey dataset available via the ESDS

The screenshot displays the 'SURVEY QUESTION BANK' interface. At the top, there is a blue button labeled 'QUESTION BANK'. Below this, the question text is shown: 'Generally speaking, do you think there is a lot of prejudice in the UK against disabled people in general, a little, hardly any or none?'. The survey title is 'Northern Ireland Life And Times Survey, 2009' and the country is 'Northern Ireland'. There are three main navigation options, each with a blue 'QUESTION BANK' button: 1) 'View response frequencies for this question' (with the Nesstar logo below), 2) 'Access full survey dataset' (with the ESDS logo below), and 3) 'View this question in other years of the same survey series'. At the bottom, there is a link for 'View this question in original questionnaire(s)' with a 'Beta' label and another 'QUESTION BANK' button.

RELU Knowledge Portal

Project: Social and environmental inequalities in rural areas

Meg Huby
University of York
01/08/2007 - 31/07/2009

RES-229-25-0004

Patterns of inequality in the distribution of social, economic and environmental goods and services in rural areas are examined. How do methods for measuring inequality differ within the natural and social sciences? And how can such differences be resolved and a common approach found? What are the implications of inequalities? Can they be regarded as unfair? And what are local residents' perceptions of local inequality and injustice.

Data created

[Social and Environmental Inequalities in Rural Areas \(SEIRA\)](#)
[P-GIS interview data](#)

Research outputs

[Article \(3\)](#)
[Paper \(1\)](#)
[Report \(1\)](#)
[Research summary \(1\)](#)
[Website \(1\)](#)

Project: Social and environmental inequality

Dataset: Social and Environmental Inequalities in Rural Areas (SEIRA)

An updated and extended version of the [SECRA spatial dataset](#) created by the research team during a previous RELU scoping study. SEIRA consist of a series of social and environmental variables for the same 6,027 rural Lower Super Output Areas in England. The SEIRA dataset classifies rural areas according to socio-economic and environmental conditions, describing patterns of inequalities in the distribution of social, economic and environmental goods and services in rural areas.

Data type: Quantitative
Data format: MS Access; CSV
Sample size: 32 social and environmental variables; 50 interviews
Country: UK
Geography: England



...ing both socio-economic and
... areas in England;
...ulties in working with spatial and
...mental science domains;
...ata for rural areas;
...t for classifying rural areas according to socio-economic and environmental conditions
...o provide sampling frames for further research and to inform policies for sustainable
...verage to Scotland and Northern
...res and output design processes.

The SECRA dataset has been compiled at the level of the new Super Output Areas (SOAs) for England. The rural extent has been identified from the new Office of the Deputy Prime Minister (ODPM) definition of urban and rural areas which relies primarily on the morphology and context of settlements.

Further information and documentation for this study may be found through the [RELU Knowledge Portal: Developing spatial data for the classification of rural areas](#).

Improvements:

- Ideas...
- Use metadata to link data/data services
- Cross-referenced publications and citations database
- Improved thesaurus interface
- Controlled vocabulary for outputs
- OAI records
- Single, umbrella search/browse interface for data, metadata, questions and the website

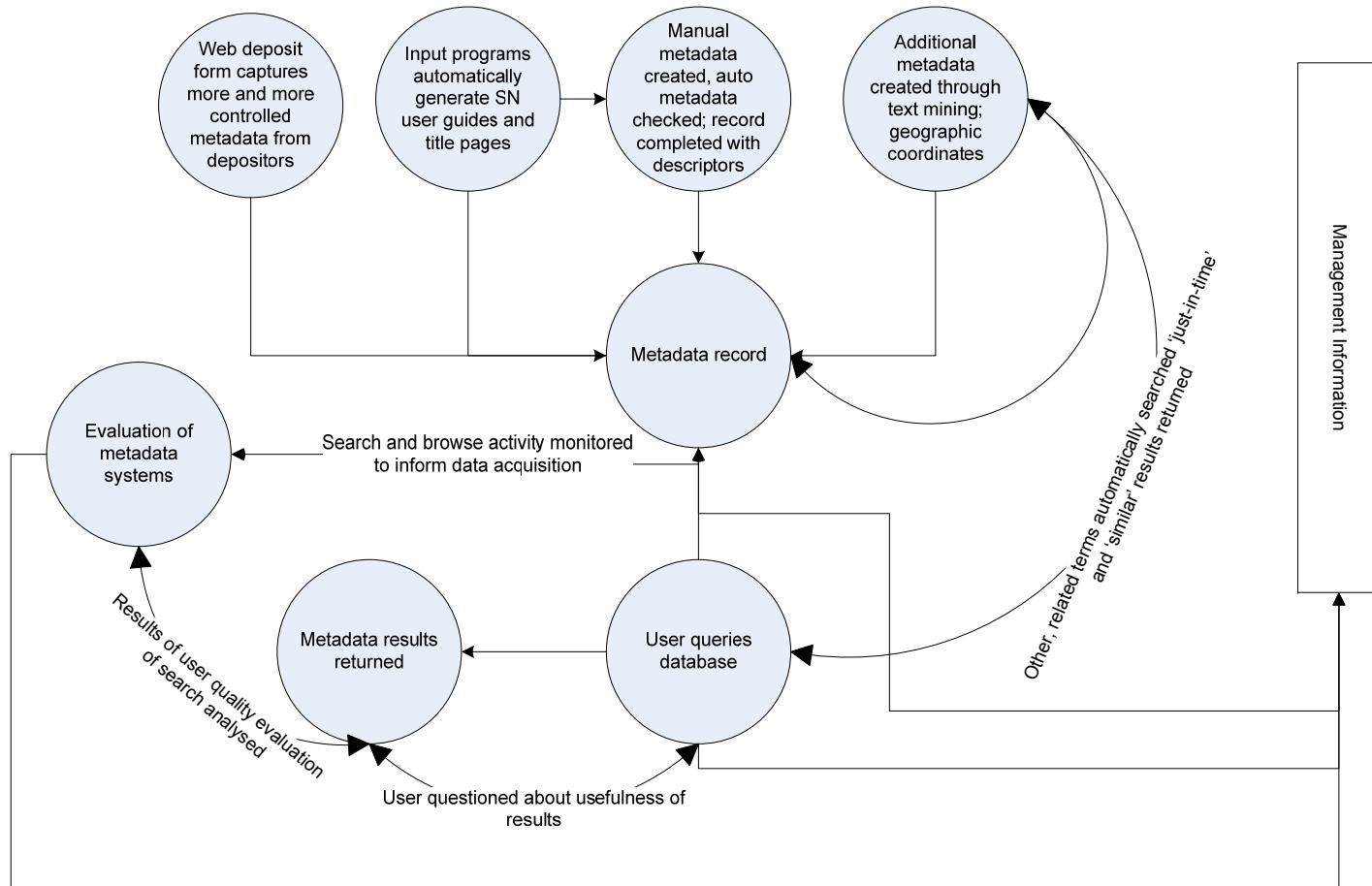
Improvements: single search interface

- Some of the options available to ESDS could be:
 - Browse to/limit the search by result type (**faceted browsing**):
 - international / qualitative / longitudinal / government / census / historical / environmental (etc) data
 - variables
 - research outputs and publications citing this article
 - question text
 - guidance
 - even news and related events
 - Go straight to the **data**:
 - visualise data (geographically, aggregated or by displaying qualitative material)
 - order data
 - download data
 - Apply **actions** on the results:
 - change the display (e.g., view full record)
 - email reference(s)/save to 'the cloud'
 - download references in MARC21/to Endnote
 - save search

In order to achieve this (longer term)...

- Richer manual metadata
 - Gather more metadata from depositors
 - Use metadata to make connections
- Increased Automatic Metadata Generation:
 - Harvesting from other websites (ESRC, ONS etc.)
 - Text mining/geographic coordinates
 - Collaborative filtering/mini-surveys
 - Relationship analysis
 - Just-in-time indexing of searches/results/user backgrounds to present more data that may be of interest

Richer manual/automatic metadata generation



In order to achieve this (more immediate)...

- Review and development of the underlying infrastructure
- All within the context of a move to DDI 3 requiring:
 - Analysis and mapping of legacy processes; proof of concepts; identification of new elements
 - Specification of a single metadata repository for the Archive, with the publication of a Full Metadata Profile
 - Platform and application development
- Timescales:
 - prioritisation will be necessary
 - longer-, medium- and short-term tasks
 - pilot project using SDS

The wider landscape: questions

- How would this vision fit into the wider social science data landscape?
- Is this the right way to deal with research outputs?
- Would such inter-connections, linking data, suit the needs of researchers?
- If so, what further connections could be made?
- What should the priorities be for improving services this way?

- Can we learn from Google and, indeed, should we?

References

- Brown S, et al. *RePAH: a user requirements analysis for portals in the arts and humanities*. Leicester: De Montfort University, 2006.
- Connaway L S and Dickey T J. *The digital information seeker: report of findings from selected OCLC, RIN and JISC user behaviour projects*. HEFCE: London, 2010.
- Hillmann D I, Phipps J. *Improving metadata quality: augmentation, recombination, services*. University of Strathclyde: Strathclyde, 2005. (part of the JISC-funded project to investigate Metadata Workflows: http://mwi.cdlr.strath.ac.uk/Documents/Improving_Metadata%20Quality2.html)
- Intrallect Ltd. *Automatic metadata generation: use cases and tools/priorities*. Intrallect Ltd (for JISC): 2009. (<http://www.intrallect.com/wiki/index.php/AMG-UC>)
- Proctor R, Williams R and Stewart J. *If you build it, will they come? How researchers perceive and use Web 2.0: a Research Information Network report*. Research Information Network: London, 2010.
- Data Documentation Initiative. *DDI 3.0 Proof of concept*. (<http://www.ddialliance.org/specification/proof-of-concept>)



CONTACT

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