

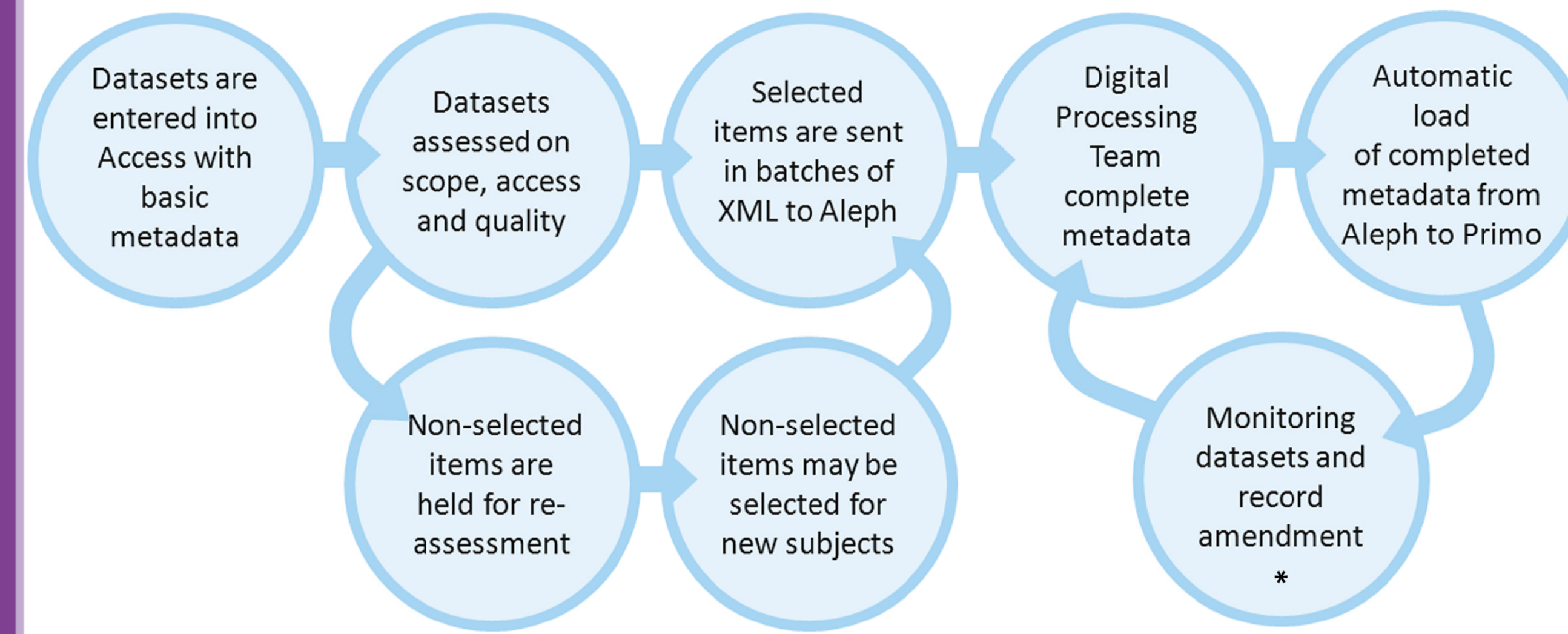
Experiments in data discovery: adapting library practices

Rachael Kotarski, Jan Ashton, Anna Gyngell and Elizabeth Newbold
The British Library, contact: rachael.kotarski@bl.uk

Background

Adopting an evidence-based approach of user needs, we developed a pilot to enable discovery of disparately held research datasets along side British Library content. User testing confirmed the suitability of the approach to meet user needs and expectations.

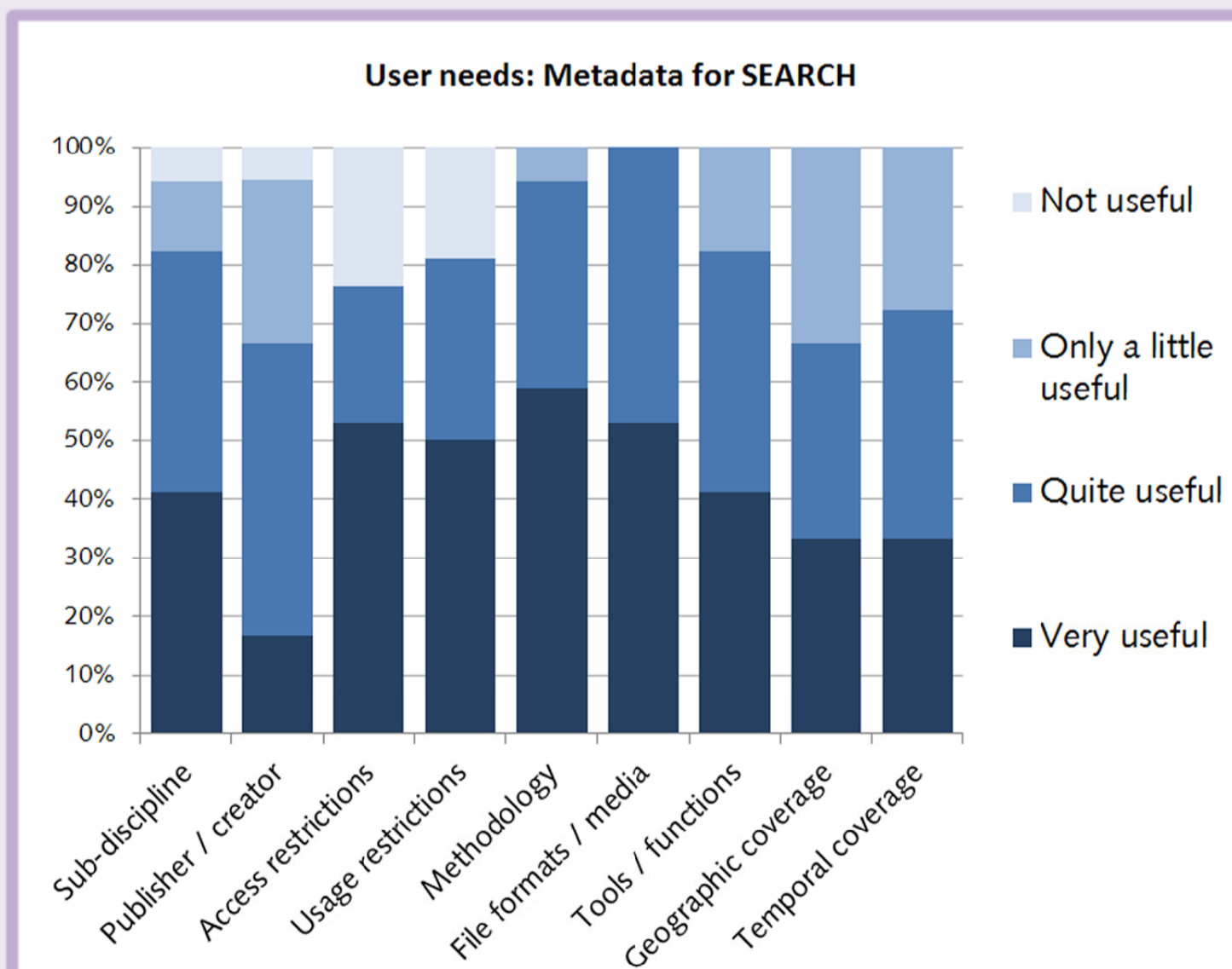
To move the pilot to a long term service, ensure its sustainability and longevity, and improve discovery with controlled terms and names, we needed to assess the feasibility of cataloguing datasets into the Library's Integrated Library System (ILS), based in Aleph. There were three steps to this assessment:



1. Mapping Metadata to MARC

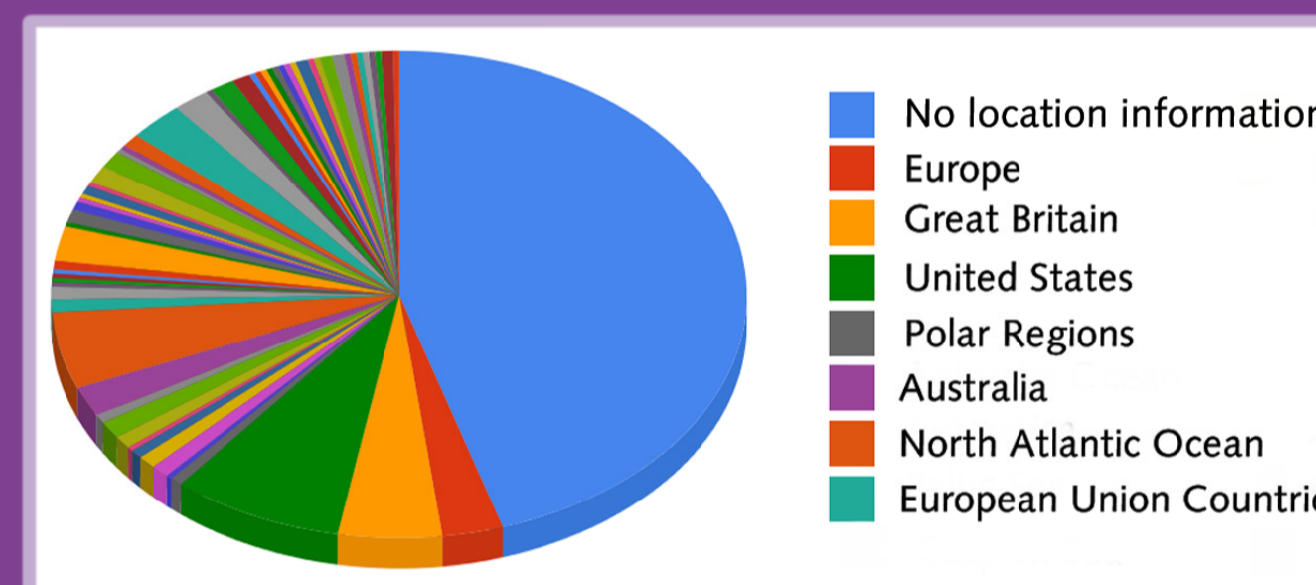
Mapping the initial Dublin Core fields to MARC and cataloguing according to AACR2 involved fields that our digital cataloguing team weren't familiar with, such as geographical and temporal coverage (of the data).

This process had to confirm whether interpretation of MARC fields could include this information.



A user survey during the pilot confirmed the metadata that users need when searching for and evaluating data, confirming the fields we needed to keep in mapping to MARC.

Name	Data
0247	10.1594/PANGAEA.120919 2 doi
24500	Partial pressure of CO2 in air, along ship track of cruise TT007 h [electronic resource]
260	[Germany] : b PANGAEA, c 2003.
538	Mode of access: World Wide Web.
500	Title from home page (viewed on Sept. 29, 2011).
500	Data is static.
516	text/plain; text/html.
520	Data was collected as part of the Joint Global Ocean Flux Study (JGOFS). Measurements of surface seawater carbon dioxide partial pressure (pCO2) in wet air were continuously monitored using an NDIR analyzer and performed according to the recommended protocols described in the 1991 DOE handbook of methods for the analysis of the various parameters of the carbon dioxide system in sea water.
506	Data are free to access.
518	1992-02-21T14:02:00 - 1992-02-26T20:34:00.
522	-140.5330, -139.6990, -0.0448, 1.4567.
540	Not stated.
584	Data were collected by the creator.
584	Accepting data submissions: No.
650 0	Seawater x Carbon dioxide content v Databases.
650 0	Ocean-atmosphere interaction z Pacific Ocean v Databases.
650 0	Chemical oceanography z Pacific Ocean v Databases.
6532	Living With Environmental Change
655 7	Dataset. 2 local
655 7	Monitoring data. 2 local
6907	Research Datasets Collection 2 bicoll
7001	Goyet, Catherine.
7102	Alfred-Wegener-Institut für Polar- und Meeresforschung.
7102	MARUM Center for Marine Environmental Sciences.
7102	Joint Global Ocean Flux Study.
85640	http://dx.doi.org/10.1594/PANGAEA.120919 y Research Datasets Collection



From top-bottom, left-right: The new workflow for selection and cataloguing of datasets; dataset MARC record, admin info removed and 'new' fields highlighted; ViewShare visualisation showing over half the dataset records have spatial coverage, a field new to cataloguers; screenshot of Explore the British Library dataset record. LEFT: Results of regular URL checks result in updates ranging from an updated URL to creation of a new record. This represents changes to 19% of all records selected.

2. Cataloguing Dublin Core records into ILS

A small number of Dublin Core records were manually keyed into the ILS to test and fine tune the mapping and ensure records looked sensible.

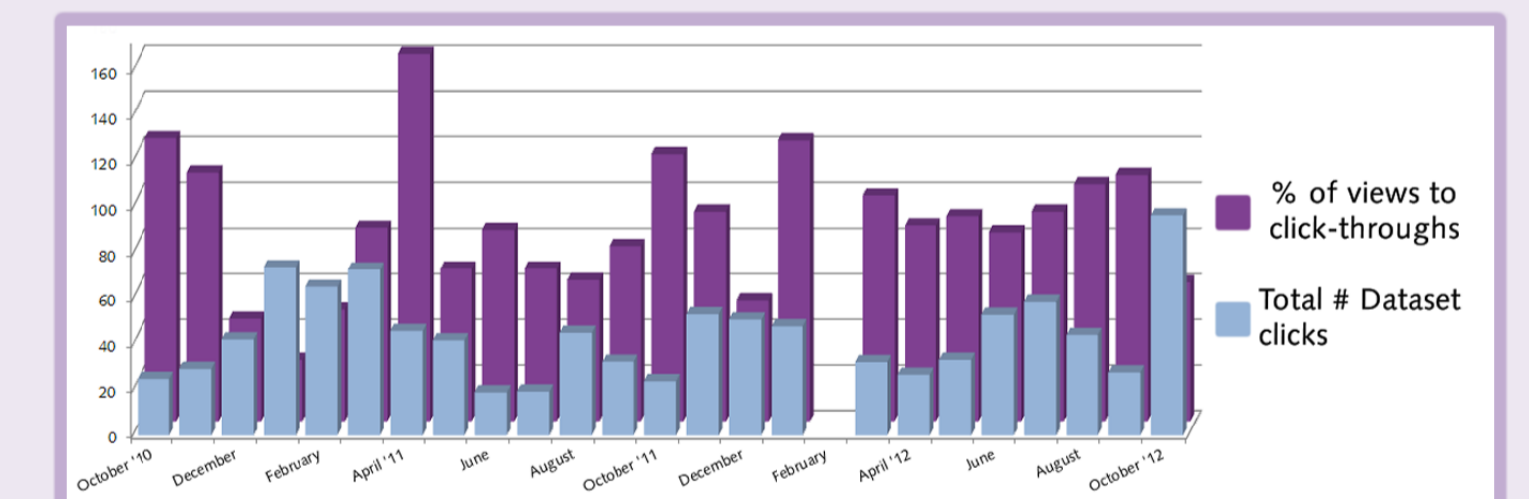
Subject experts supported cataloguers in understanding fields that were new to them and interpreting and incorporating them into AACR2 rules. Subject experts also required support from cataloguers, to understand the AACR2 requirements and how these related to what users would see in each record.

3. Cataloguing new dataset records into the ILS

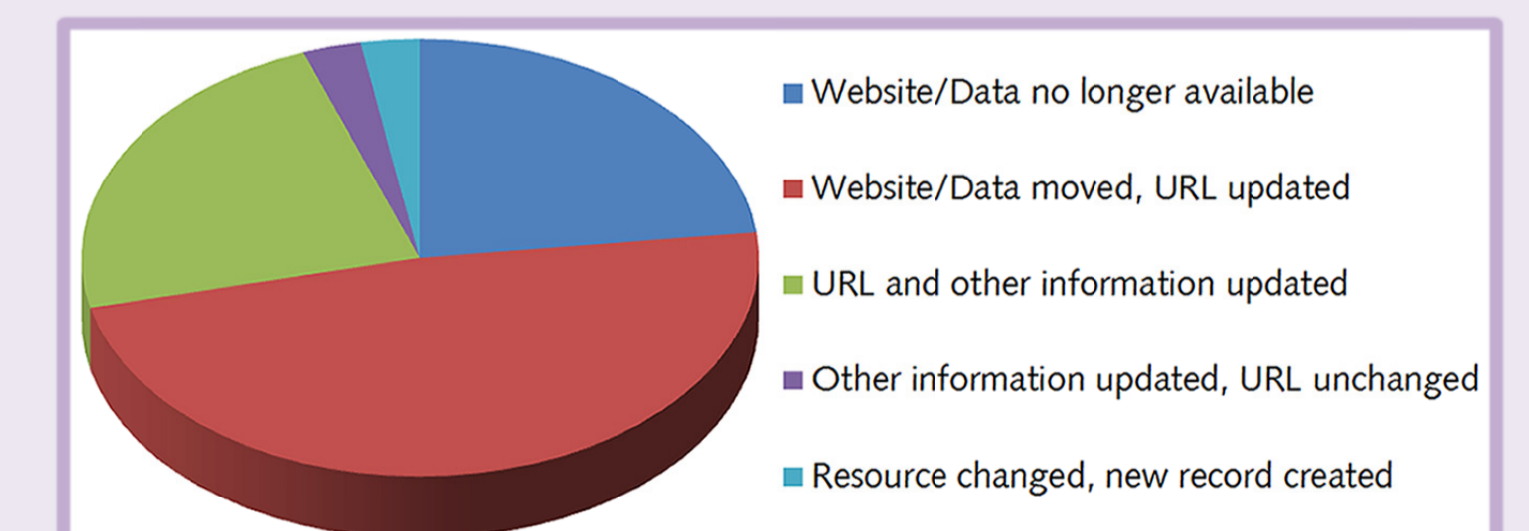
Finally cataloguers entered completely new dataset records into the ILS without subject support.

As well as demonstrating the new process would be practical, this allowed us to test the workflow from selection to cataloguing, and prove that cataloguing datasets is within the current expertise of cataloguers.

Usage is a factor in success, and despite the small pilot size, has remained consistent over time with users going on to view datasets.



* Datasets may change or disappear without notice, so our next challenge is maintaining accurate records.



500 datasets on environmental change, biodiversity and neglected tropical diseases are now catalogued and available via explore.bl.uk. The continuing aim will be to monitor sustainability of the service and investigate metadata harvesting from third parties, including DataCite.

This work shows that it is possible to use current library systems to

provide discovery for externally held content and datasets, and current library practices can be adapted to meet that goal. The expertise and experience of current staff is still well matched to the aim of providing discovery of research datasets alongside traditional library content without the need for a separate tool to do so.

<http://bl.uk/science>
<http://bl.uk/datasets>
<http://bl.uk/bibliographic/main.html>

See ViewShare record via QR code:

