

CEH LANCASTER
LANCASTER ENVIRONMENT CENTRE, LIBRARY AVENUE,
BAILRIGG, LANCASTER, LANCASHIRE LA1 4AP

**A STUDY TO EXAMINE THE OPERATION AND FUNCTION
OF A VIRTUAL UK ENVIRONMENTAL SPECIMEN BANK
(UK-ESB)**

FINAL REPORT

J.S. Chaplow, S. Patel, D.C. Howard, M.J. Brown and R.F. Shore¹

Project Leader: Richard F. Shore
Contract Start Date: 1st November 2011
Report Date: 30th September 2012
Report To: UK-EOF
CEH Project No: NEC04510
CEH Report Ref No: LA/ NEC04510/1

This is an unpublished report and should not be cited without permission.

¹ Corresponding author

INTELLECTUAL PROPERTY RIGHTS

CONFIDENTIALITY STATEMENT

'In accordance with our normal practice, this report is for the use only of the party to whom it is addressed, and no responsibility is accepted to any third party for the whole or any part of its contents. Neither the whole nor any part of this report or any reference thereto may be included in any published document, circular or statement, nor published or referred to in any way without our written approval of the form and context in which it may appear'

1 Executive Summary

1. Environmental specimen banking is recognised internationally as an integral part of long-term environmental research and monitoring. Analysis of preserved environmental samples is often needed to detect and quantify patterns and rate of environmental change, and the emergence and progression of environmental hazards and risks.
2. National Environmental Specimen Banks have been established in several countries; they vary in scope and breadth. There are a few specialised environmental specimen holdings in the UK but no national-scale catalogue of holdings, despite an estimated annual spend of £16 million to store specimens. This lack of information results in under-exploitation of archived specimens and is a lost opportunity to facilitate world-class science and identify emerging pressures and threats on the environment.
3. An earlier project had identified key stakeholder organisations either engaged in archiving nationally important environmental specimens or who wished to utilise such specimens. These stakeholders had agreed there was a need for a national metadata catalogue of environmental specimens (subsequently termed *a virtual UK-ESB*). The objective of the current project was to further develop a virtual UK-ESB. Specifically, the aim was to work with stakeholders to establish the correct metadata entry fields, the search capabilities, the functionality and the nature of the hosting website of a virtual UK-ESB.
4. More than 80 stakeholder organisations that had previously expressed an interest in a UK-ESB were approached to provide feedback either electronically or by attending a stakeholder workshop. Thirty eight organisations responded. All remained interested in the UK-ESB concept and seventeen answered the survey questions.
5. Mock-ups of data entry screens, search screens and ideas around the functionality of a UK-ESB were developed by the CEH project team. These were mailed to stakeholders for feedback. Initial feedback was incorporated into the mock-ups which were then presented for discussion at a workshop comprising 15 attendees from across the specimen archiving community, CEH and the UK-EOF.
6. Workshop participants reviewed and agreed the format of 23 mandatory or optional data-entry fields for a virtual UK-ESB that, in the absence of standard for material samples and archives, were aligned with ISO19115 (geospatial metadata standard) and DublinCore (metadata standard). These fields were sub-divided into the following headings: *Sample description*, *Categorisation*, *keywords and links*, *Storage Information* and *Contact information*. Workshop participants also made a number of recommendations as to the format of the data entry screens and inclusion of extra fields.

7. Workshop participants reviewed options for search capabilities and made recommendations as to simple and advanced searching methods and their formats. It was also recommended that search facilities of the ESBs of other countries be examined to determine what is used, ease of use, and how they match the recommendations from the workshop.
8. Workshop participants reviewed options for functionality and agreed a detailed list of prioritised requirements.
9. Workshop participants agreed that a virtual UK-ESB should be hosted through a dedicated website that would also provide wider information, such as recently updated or added specimen holdings, most downloaded information, links to other groups, standard operating procedures, etc.
10. The next step for the development of a virtual UK-ESB is to implement the design and development ideas captured in the current report and build a test version of a virtual UK-ESB. This would be tested and refined, and could then be launched on a specifically designed website. This would need to be accompanied by a communication strategy. There is potential to link and co-brand a virtual UK-ESB with the UK-Environmental Observation Framework (UK-EOF).

Table of Contents

1	Executive Summary	1
2	Introduction	4
2.1	The concepts and benefits of Environmental Specimen Banks	4
2.2	Current national Environmental Specimen Banks in other countries	4
2.3	Specific specimen banks currently in existence in the UK	5
2.4	The need for a UK-ESB	5
2.5	Progress towards the development of a UK-ESB	6
2.6	A virtual UK-ESB	6
2.7	Aims of the current project in developing a virtual UK-ESB	7
3	Description of project methods	8
3.1	Initial contacts with stakeholders	8
3.2	Preparations for the workshop.....	9
3.3	Compliance with Standards.....	10
3.4	Stakeholder workshop	11
4	Draft virtual UK-ESB design and functionality	12
4.1	Data entry fields	12
4.2	Search capability	14
4.3	UK-ESB functionality	16
4.4	Web portal.....	17
5	Conclusions and next steps	18
6	Appendices	19
	Appendix 6.1. List of those that were willing to look at and comment on mock-up UK-ESB catalogue screens.	19
	Appendix 6.2. Screen mock-ups and associated fields for virtual UK-ESB data entry screens	21
	Appendix 6.3. MuShCoW for data entry screens and fields.....	26
	Appendix 6.4. MuShCoW for functionality.....	27
	Appendix 6.5. MuShCoW for web front end	28

Appendix 6.6. Updated list of fields (with explanation) that incorporates feedback from interested parties and attendees of 2012 workshop.....28

2 Introduction

2.1 The concepts and benefits of Environmental Specimen Banks

Long-term collection and preservation of environmental specimens (often termed environmental specimen or sample banks – ESBs) allows analysis and evaluation of samples both as a part of real-time monitoring and basic research. ESBs also enable investigators to extend their research into the past. Such analyses are often needed to detect and characterise patterns and rate of environmental change and the emergence and progression of environmental hazards and risks. Examples include quantification of changes in species dietary patterns and migration (through stable isotope analysis), identification of the presence and rate of spread of emerging contaminants, pathogens and diseases, and measurement of genetic drift in species. One example of the scientific value and excellence of ESBs is demonstrated by the Japanese Environmental Specimen Bank for Global Monitoring (es-BANK) which has gained 14th place in citation ranking among academic publications on environmental sciences. The wider impact of ESBs is that they enable detection and characterisation of long-term changes in response to environmental drivers. The importance of this is evident from our pressing need to understand environmental resilience to global climate change.

2.2 Current national Environmental Specimen Banks in other countries

According to the United Nations Environment Programme (UNEP) Production and Consumption Unit, specimen or sample banks have been established in several countries both for environmental and human samples. The countries that currently operate some form of sample bank include Germany, Japan, Nordic Countries, Italy, Canada and the USA; others are under development elsewhere.

The scope and breadth of the current existing national specimen banks vary between countries. Some ESBs have a wide remit. For example, the National Wildlife Specimen Bank in Canada archives samples from a wide range of Canadian habitats while the Japanese es-BANK archives samples from all over the world. Other ESBs are smaller and act as centralised repositories for samples from specific ecosystems. For instance, the U.S. Marine Environmental Specimen Bank cryogenically banks well-documented environmental specimens (such as marine mammal tissues, mussels, oysters, fish tissues, seabird eggs) collected as part of US marine research and monitoring programs. The German ESB, the Umwelt Probenbank, only accepts and stores limnetic, marine, terrestrial and human samples collected from designated sampling areas on an annual basis, while the Paljakka Environmental Specimen Bank (Finland) is restricted to moss, lichen and bark samples.

The International Environmental Specimen Bank Group (IESB) promotes the world-wide development of techniques and strategies for environmental specimen banking. There is a breadth of information available from their website² including details of existing specimen banks, standard operating procedures and protocols, previous meetings and their proceedings.

2.3 Specific specimen banks currently in existence in the UK

A number of specialised environmental and human specimen catalogues are in existence in the UK of which we are aware and which are searchable through some form of electronic functionality. These include:

- British Bryological Society UK Catalogue of specimens – a Microsoft Excel sheet with more than 41000 rows that exists to promote the study of liverworts and mosses³
- Natural History Museum (NHM) collections including botany, entomology, mineralogy, zoology and palaeontology e.g. *type specimens of birds in the NHM* – a searchable electronic list of types of birds in the museum intended for use by the scientific research community, *Zoology Specimen catalogue (NHM)* – a searchable database with 21 possible fields. According to the home page as little as 10% of the 28,000 samples have so far been catalogued⁴
- University of Cambridge University Museum of Zoology – a searchable online main catalogue and draft bivalve catalogue⁵
- UK Biobank – holding for future analysis samples of blood, urine and saliva samples from half a million people aged 40 - 69 (samples taken 2006-10) and whose long term health outcomes are monitored⁶
- SurgiCat – the Royal College of Surgeons of England. The catalogue is searchable and contains some environmental specimens⁷

2.4 The need for a UK-ESB

Formal environmental specimen banking is recognized internationally as an integral part of long-term environmental research and monitoring. There are numerous monitoring and experimental research studies in the UK that involve collection and archiving of environmental specimens. The United Kingdom Environmental Observation Framework (UK-EOF), part of the Living With Environmental Change (LWEC) partnership, catalogues environmental observations made for and by the UK, but there is no equivalent catalogue for

² <http://www.inter-esb.org>

³ <http://www.museumwales.ac.uk/en/3109/>

⁴ <http://www.nhm.ac.uk/nature-online/collections-at-the-museum/>

⁵ <http://www.museum.zoo.cam.ac.uk/collections.archives/catalogues/>

⁶ <http://www.ukbiobank.ac.uk/>

⁷ <http://surgicat.rcseng.ac.uk>

UK environmental specimen holdings. The only information available is with regard to specific collections, such as those listed above.

As part of a previous study in 2009/10, we estimated the cost of UK projects that store or archive specimens to be approximately £16 million per year (Chaplow et al. 2010)⁸. The lack of any metadata catalogue for these holdings means that it is impossible to readily discover what specimens have been collected, who curates them, and whether they are accessible for research purposes. Current UK collections of environmental specimens are almost certainly under-exploited because of a general lack of awareness about their existence. This represents a lost opportunity on the part of the UK to facilitate world-class environmental science and to help government and regulatory agencies identify emerging threats to, and pressures on, the environment.

2.5 Progress towards the development of a UK-ESB

The Centre for Ecology & Hydrology (CEH) held a workshop in 2010 that examined the need for, and desirability of, some form of UK Environmental Specimen Bank (UK-ESB). This workshop involved key stakeholder organisations who were engaged in collecting and archiving nationally important environmental specimens for various purposes, such as education, the promotion of knowledge, and scientific use. Other stakeholders were organisations who utilise, or would wish to utilise, such specimens.

The workshop established that there was general acceptance and enthusiasm from the stakeholder community to engage in a move towards a UK-ESB. A strategic development plan was formulated (Chaplow et al. 2010). The agreed key first step was the establishment of a *virtual* UK-ESB.

2.6 A virtual UK-ESB

A virtual UK-ESB would essentially be a standardised metadata catalogue, presented through a web portal with an associated website. The catalogue would adopt consistent descriptions of specimen archives. The catalogue would also provide links to each of the various individual sample archives, thereby providing awareness of and access to samples and sample holders.

A virtual UK-ESB would require a national partnership between holders of UK environmental specimens. The partners would enter information into the virtual UK-ESB about their specimen holdings but no changes to current archiving practices, ownership or intellectual property rights would be required. The virtual UK-ESB would be expected to yield significant benefits to specimen holders. These could include sharing of data, samples

⁸ Chaplow, J.S., Walker, L.A., MacKechnie, C.J. & Shore, R.F. 2010. *A scoping study of specimen archiving activity in the UK and the potential for a UK Environmental Specimen Bank*. Centre for Ecology & Hydrology report to the Environmental Research Funders Forum. Centre for Ecology & Hydrology, UK, pp. 26, ISBN: 978-1-906698-17-1. <http://www.erff.org.uk/documents/201005-ceh-scopingstudy.pdf>

and facilities, promotion of best practice and facilitation of strategic links with other types of specimen banks (human, DNA banks etc). For individuals or organisations that wanted to access archived environmental specimens, a virtual UK-ESB would be a readily accessible and searchable web portal where they could discover information about what UK specimen holdings exist and how they might be accessed.

Overall, the aim of a virtual UK-ESB would be to:

- improve the discoverability of, and access to, physical samples to help maximise the benefits gained from the current UK investment in archiving environmental specimens. This will improve the ability of UK researchers and agencies to address pressing environmental issues

2.7 Aims of the current project in developing a virtual UK-ESB

The overall aim of the current project was to further develop the concept of a virtual UK-ESB. Specifically, the aim was to develop agreement, through work with stakeholders, on the following four areas of the virtual UK-ESB.

- a) *Metadata fields for the virtual-ESB.* This involved review of an extensive list based on initial ideas collated at the 2010 CEH workshop (Chaplow et al., 2010). The aim was to reduce and make a final selection of metadata fields such that they matched data availability, ease of capture, relevance and data standards (existing and under development)
- b) *Search capability.* A key value of a catalogue is in its interrogation. The virtual UK-ESB will be used by a variety of parties and each is likely to ask a different style of question. To develop the appropriate functions, the questions being asked, the material being held and agreement of the data sources to provide answers must be combined.
- c) *Functionality.* Interaction with holders of specimen collections will be key to the size and quality of the catalogue. Not only do they have to understand and approve of the information held in the catalogue, but their interaction with the catalogue must not deter users from entering data.
- d) *Web portal.* An agreed structure is needed for the web portal so that it provides information effectively and is a means of easy access to the metadata catalogue.

An outline of the chronology of the project work is described briefly in section 3 and the work on developing the four areas of the virtual UK-ESB are described in section 4.

3 Description of project methods

3.1 Initial contacts with stakeholders

The overall approach was to develop the outputs of the CEH 2010 workshop, obtain information and feedback through electronic contact, and to test mock-ups and ideas on metadata fields, interface, functionality and web portal, through a targeted workshop with key stakeholders.

Chaplow et al. (2010) originally identified more than 80 parties that might be interested in a UK-ESB. These were re-contacted to confirm their interest, to help identify other interested parties that has been missed originally, and to canvass them for feedback on UK-ESB draft designs. Those who expressed an interest (see Appendix Table 6.1) were invited to a one day workshop in April 2012.

In the initial re-contact with the original stakeholders, the following questions were asked:

- Are you willing to be involved in the next phase of a UK-ESB? If not, is there someone else in your organisation we should contact?
- Are you willing to comment on a mock up catalogue for samples and archives?
- Who uses your samples at the moment and do you receive requests for access to samples/archives?
- Can you think of others who may be interested in using your samples or material?
- Are you able to attend a workshop in April 2012 to discuss and develop the concept of a UKESB further?

Previously un-contacted potential users of samples or materials that were not also collectors were asked the following additional questions:

1. Do you use environmental samples as part of your work?
2. If samples were available, would you use them?
3. Are there others that we should contact within your organisation who collect or use samples and archived material?

There were 38 responses received to the initial contact approach. All respondents were still interested in the concept of a UK-ESB and were willing to examine and comment on mocked- up screens for data entry into a catalogue; seventeen answered the survey questions. These responses included detail around who currently requires access to specimens, archives and collections and listed others who may be interested in the initiative.

Respondents not currently collecting samples were interested in knowing what samples are being collected and archived, where they are stored, whether they are accessible and what they had been used for. Fifteen respondents indicated that they were available to attend a workshop.

3.2 Preparations for the workshop

Before the workshop was conducted, the following material was prepared:

a) *Mock-up screens for data entry*

There is currently no standard available for material samples and archives but there are for electronic scientific data. For that reason, the mock-ups were designed to be aligned with ISO19115 (geospatial metadata standard) and DublinCore (metadata standard). DublinCore has 15 core fields and ISO has 17 core fields although some of the fields overlap.

A list of 33 possible data entry fields that were suggested in the 2010 CEH workshop (Chaplow et al. 2010) was reviewed. An iterative series of internal discussions within the CEH project team was used to refine this list and reduce it to 23. Some of these were considered to be essential or necessary to be compliant with metadata standards, and so were mandatory fields, others were considered useful but not essential, and so were to included as non-mandatory fields (labelled (O) for Optional in the mock ups).

Where possible, the names of each data entry field were drawn from and were in alignment with existing metadata standards. In addition, an alternative title was provided since standardised names are sometimes difficult to decipher, and an explanation on the detail required was included.

The data entry fields were converted into four mock-up data entry screens: *Sample description; Categorisation, keywords and link;*, *Storage Information; Contact information*. These were designed to be “straw men” for comment in the workshop. No symbology, colour or imagery was used on the screens so as to avoid distractions when they were presented at the planned workshop.

b) *Mock-up search screen*

A single simple search mock-up screen was generated for presentation at the workshop. Examples of searches that people have used and found useful were requested. The screen contained a free search box (in the style of Google) and a more structured filter search (in the style of Web of Science).

c) *A demonstration of the CEH Information Gateway data catalogue*⁹

This was prepared to demonstrate the kind of functionality a catalogue may have.

d) *Wider availability of presentation material*

The mock-ups and presentations used in the workshop were made available (via a wiki¹⁰) to interested parties who were unable to attend the workshop.

⁹ <https://gateway.ceh.ac.uk/>

¹⁰ https://wiki.ceh.ac.uk/display/ukesb/description_entry

3.3 Compliance with Standards.

For this study the catalogue fields were selected in broad alignment with ISO19115 (geospatial metadata standard) and DublinCore (metadata standard). These standards were selected in liaison with technical experts at CEH who set up the CEH Information Gateway data catalogue. The Gateway is a searchable on-line tool that provides metadata (information about data), data download and display functions (map viewer). The Gateway was designed to be aligned with the INSPIRE standard which was set up to establish an infrastructure for spatial information in Europe. Working in compliance with standards enables catalogues to communicate and provides the added benefit of inter-operability. For example, the Gateway broadly complies with the INSPIRE directive and the ISO19115 standard, therefore information recorded in the Gateway is available to other INSPIRE/ ISO19115 compliant portals e.g. data.gov.uk and Joint Research Centre (EU Commission) Web Mapping Services¹¹ (WMS).

A number of suggestions for other standards were received from interested parties. These standards were examined and compared with ISO19115 and DublinCore to ensure that the UK-ESB would accommodate samples from a wide range of archiving organisations.

The International Council on Archives Committee on Descriptive Standards (ICA/CDS) adopted the General International Standard Archival Description (ISAD(G))¹² at a meeting in Canada in 1999. The description was designed to be used in conjunction with existing national standards or as the basis for the development of national standards and to identify and explain the context and content of archival material in order to promote its accessibility.

The concept and need for standards associated with archival description was discussed by the ICA at their first meeting in 1988. A wealth of information is available from the ISAD(G) and from the website of the ICA/CDS¹³ including proceeds of meetings, reports and the standard and guideline documents. The standard divides the description into seven elements: identity statement area, context area, content and structure area, conditions of access and use area, allied materials area, and description control area. These elements can be broadly related to the fields identified in a proposed UKESB catalogue which was designed to be aligned with ISO19115 and DublinCore.

The Collections Trust¹⁴ is an independent UK based charity working to help helping cultural organisations to unlock the potential in their collections. The Trust has an Intelligence Hub entitled the Collections Link¹⁵ that provides the Spectrum Standard¹⁶ which is used by more than 7,000 museums, galleries and cultural organisations worldwide. Registered users can download a licensed copy of the Spectrum Standard free of charge.

¹¹ http://eusoiils.jrc.ec.europa.eu/wms/WMS_Right.htm

¹² <http://www.icacds.org.uk/>

¹³ [http://www.icacds.org.uk/eng/ISAD\(G\).pdf](http://www.icacds.org.uk/eng/ISAD(G).pdf)

¹⁴ <http://www.collectionstrust.org.uk/>

¹⁵ <http://www.collectionslink.org.uk/>

¹⁶ <http://www.collectionslink.org.uk/programmes/spectrum>

The Spectrum standard was first published in 1994 after an extensive collaborative development project and has been updated several times since. The standard has eight Primary Procedures including Object entry, Acquisition, Location and movement control, Cataloguing, Object exit, Loans in, Loans out and Retrospective documentation. Version 4 of the Spectrum standard is 95 pages in total and not all of the procedures are applicable but the information should be taken into account for creation of a UK-ESB catalogue that aims to accommodate Museum collections.

3.4 Stakeholder workshop

A workshop with 14 attendees (see Table 1.) from across the specimen archiving community (including CEH and UK-EOF) was held on April 27th, 2012. Attendees were provided with background information regarding a virtual UK-ESB initiative. The workshop was divided into 3 main discussion areas: catalogue data entry, catalogue search capabilities and overall functionality. In addition, options for the web portal or ‘front end’ of the catalogue were discussed. Where mock-up screens were used, workshop attendees were asked whether the fields were appropriate for their data and whether the screens were suitable for their use.

Comments and questions from workshop attendees were collated and categorised into one of the four categories: MuShCoW - **M**ust have, **S**hould have, **C**ould have and **W**on’t have in which all suggestions were recorded (see Appendices B-D in Section 6).

Development ideas and feedback from the workshop on metadata fields, search function, functionality and web portal are presented in the next section of this report.

Table 1. Attendees of a workshop held 27th April 2012 at Birmingham Aston Court.

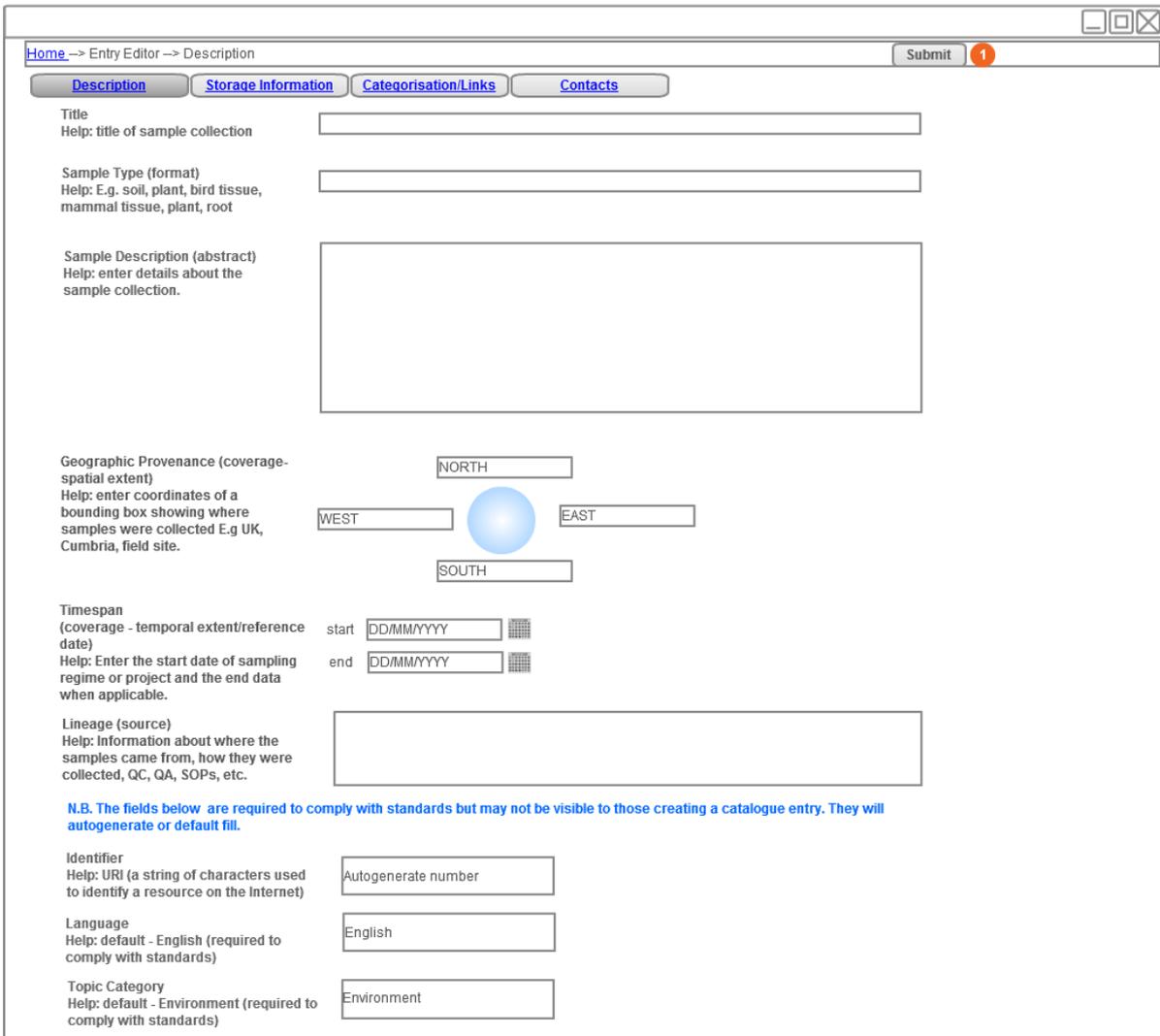
Name	Affiliation
Alexandra Tomlinson	FERA
Amber Vater	UK-EOF
Chris Shortall	Rothamsted Research
David Howard	CEH
Helen Peat	British Antarctic Survey
Jacky Chaplow	CEH
Jerry Herman	National Museums Scotland
Mike Brown	CEH
Monika Juergens	CEH (Fish Tissue Archive)
Richard Harrington	Rothamsted Research
Richard Shore	CEH
Roy Neilson	The James Hutton Institute
Sabera Patel	CEH
Zena Floody	National Museums Scotland

4 Draft virtual UK-ESB design and functionality

4.1 Data entry fields

For illustrative purposes, the mock-up screen for the *Sample description* page is shown in Figure 1 and the set of data fields (and their descriptions) that relate to this page are given in Table 2. All four mock-up screens (*Sample description*, *Categorisation*, *keywords and links*, *Storage Information* and *Contact information*) and their respective data field tables are given in Appendix 6.1 (Section 6).

The screens had no database or functionality nor did they contain any data validation, editing or status. Text blocks shown on the screens were more verbose than necessary in a web-enabled system as no help system (general or targeted) was present.



The image shows a web browser window with the address bar displaying "Home -> Entry Editor -> Description". The browser title bar includes "Submit" and a red circle with the number "1". Below the address bar are four tabs: "Description" (selected), "Storage Information", "Categorisation/Links", and "Contacts".

The main content area contains the following fields and labels:

- Title**: Help: title of sample collection. Input field.
- Sample Type (format)**: Help: E.g. soil, plant, bird tissue, mammal tissue, plant, root. Input field.
- Sample Description (abstract)**: Help: enter details about the sample collection. Large text area.
- Geographic Provenance (coverage-spatial extent)**: Help: enter coordinates of a bounding box showing where samples were collected E.g UK, Cumbria, field site. Includes input fields for NORTH, WEST, EAST, and SOUTH, and a blue globe icon.
- Timespan (coverage - temporal extent/reference date)**: Help: Enter the start date of sampling regime or project and the end data when applicable. Includes "start" and "end" labels and date input fields (DD/MM/YYYY) with calendar icons.
- Lineage (source)**: Help: Information about where the samples came from, how they were collected, QC, QA, SOPs, etc. Input field.
- Identifier**: Help: URI (a string of characters used to identify a resource on the Internet). Input field with "Autogenerate number" text.
- Language**: Help: default - English (required to comply with standards). Input field with "English" text.
- Topic Category**: Help: default - Environment (required to comply with standards). Input field with "Environment" text.

A note below the fields states: "N.B. The fields below are required to comply with standards but may not be visible to those creating a catalogue entry. They will autogenerate or default fill."

At the bottom left, a red circle with the number "1" is followed by the text: "Content will be submitted in a format which can be searched".

Figure 1. Mock up data entry screen for *Sample description* page

Table 2. Description fields, alternative names, explanations and examples for the fields used in the *Sample description* screen

<u>Field name</u>	<u>Alternative (comply with standards)</u>	<u>Help: Information on what is required</u>	<u>For example</u>
Title		Title of sample collection	Predatory Bird Monitoring Scheme Tissue Archive
Sample type	Format		soil, plant, bird tissue, mammalian tissue, plant, root, etc
Sample Description	Abstract	Enter details about the sample collection	
Geographic Provenance	Coverage - spatial extent	Enter coordinates of a bounding box showing where samples were collected	UK, Cumbria, Field site
Timespan	Coverage - temporal extent/reference date	Enter the start date of the sampling regime or project and the end date when applicable	
Lineage	Source	Information about where the samples came from, how they were collected, QC, QA, SOPs	
Identifier	URI	String of characters used to identify a resource on the internet	
Language		Default: English	
Topic category		Default: Environment	

Please note the last three fields in Table 2. These are mandatory fields, required for the virtual UK-ESB to comply with ISO19115 and Dublin Core standards. These fields are unlikely to vary from a default setting and may not be required to be visible on the data entry screen as they can be auto-generated using a default fill. Proposed contents are *Identifier* (Auto-generate: URI), *Language* (Default: English), *Topic category* (Autofill: Environment-an ISO19115 standard field).

The workshop objectives were to check for completeness, avoid confusion and misunderstanding, record views on the format and freedom of responses and gauge the level of buy in and likely response to requests for information.

Feedback from the workshop attendees was that:

- The catalogue should have a limited number of screens and the sizes of boxes should be small so that users are not over-faced by the amount of information needed
- Screen buttons should be employed for opening input text boxes
- Field names should be unambiguous and there should be dynamic help available, both in brief and in full. The help should explain the level of detail required (in terms of space e.g. 100 or 1000 characters and information)

- Duplication of data capture should be avoided; use automated prompting wherever possible. Record information in a way that is useful from a users' perspective
- Use (where applicable) different data entry types and formats (buttons, drop downs)
- Defaults and autofill should be employed allowing changes when not applicable (e.g. Language. Default English may need to be changed to Latin)
- Existing thesauri should be used; a bespoke thesaurus would be ideal, but would require too much resource to develop and maintain
- Access restrictions must be spelled out (use standard terms e.g. drop down menus)
- Users should be able to record location information in different ways (e.g. bounding box, name of place, latitude longitude)
- Cost implications of access to samples should be recorded
- Personal names of the collectors, collecting organisations and funding bodies should be entered separately from the contacts list. Format for data entry should be controlled e.g. Surname, first forename second forename.
- Additional fields were requested.
 - Resource Type (identify level of description e.g. collection, item, whole repository).
 - Size of collection (e.g. approximate number of items).
 - Information on custodial history (how and when it got into the repository rather than initial collection information).
 - Physical condition (may affect access restrictions; could trigger conservation work)
 - Include preferred wording for citation for the collection.
 - Include expeditions or projects name
 - Accrual rate or not accruing (i.e. will the collection be added to).
 - Date of original entry, by whom plus date of last update, by whom.

The MuShCoW list on data entry fields that was compiled at the workshop is given in Appendix 6.3 (Section 6).

4.2 Search capability

A single simple search screen (Figure 2) was presented to workshop attendees for discussion and examples of searches that people have used and found useful were requested. The screen contained a free search box (in the style of Google) and a more structured filter search (in the style of Web of Science).

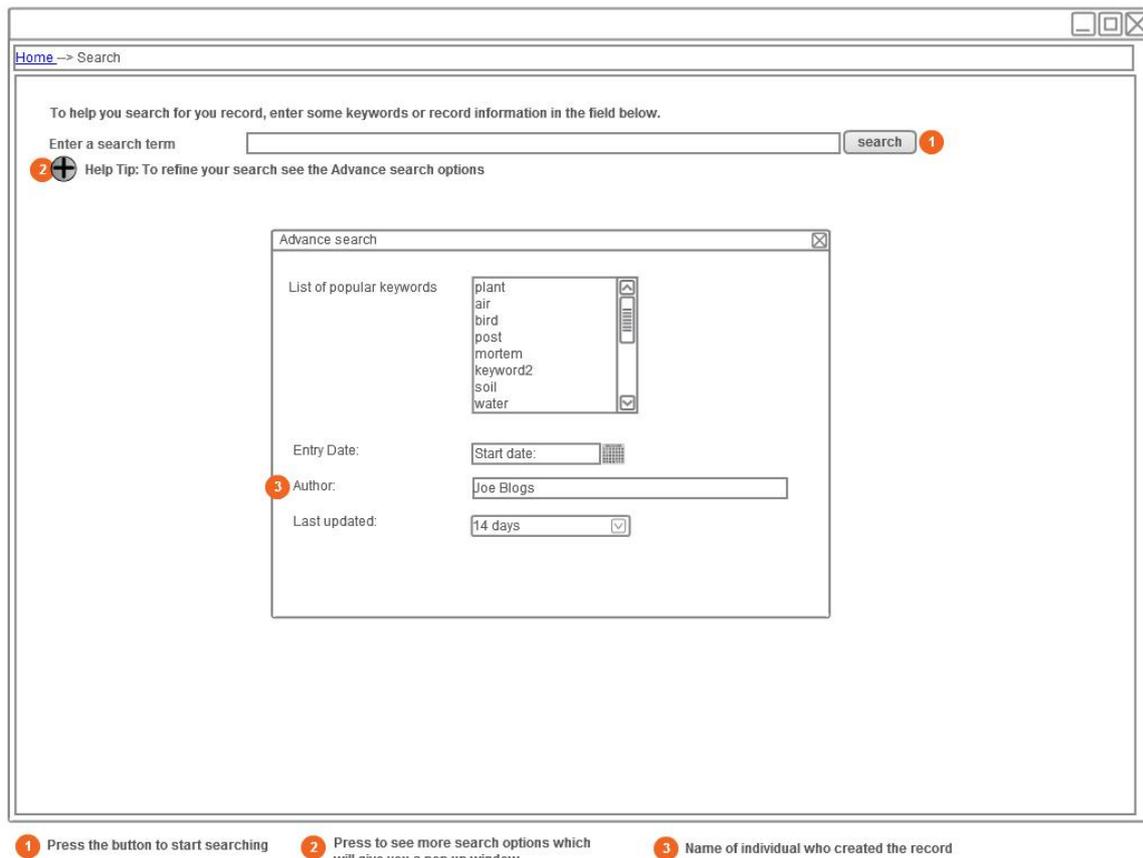


Figure 2. Simple search screen mock-up.

Workshop attendees suggested looking at the search tools available from the websites of the German ESB and other ESBs. They also thought that the search screen must identify, target and relate to users and be structured, progressive, possibly map based, with drop downs menus. In addition the search should be simple in the first instance with more detailed options available as outlined below.

- Simple (access, abstract and text, use own keywords, broad categories)
 - Hierarchical (mammal, plant, soil, etc.)
 - Boolean (in computer programming languages Boolean represents either true or false)
 - Structured or directed (e.g. sample type, geographic location, temporal, material, etc.)
- More advanced
 - Web of Science search (Boolean multiple structured)
 - Amazon type of search (general search followed by hierarchical drilling down)
- Temporal information (including date last modified or updated) could be used as a ranking.

4.3 UK-ESB functionality

Workshop attendees were shown the demonstration of the CEH Information Gateway data catalogue to demonstrate issues around catalogue functionality and explain where many of the ideas for the virtual UK-ESB had originated from.

The presentation was used to introduce questions of functionality including: authentication, deletion policy, bulk upload, export and update, publication and workflow, monitoring and reporting, auditing, search engine optimization (SEO) and feedback. Each issue was presented and discussed in turn.

Feedback from the workshop attendees was:

- The catalogue should not have a gatekeeper; data may be added by, but not be necessarily owned by a single user. Ownership may be dealt with by user groups who are able to update and edit entries.
- There should be a workflow i.e. save as draft, private, submit for publication, submitted, approved.
- Validation of data entries (i.e. checking that a record is complete) should be automatic (e.g. mandatory fields may be highlighted) although it will be possible to save incomplete entries as draft.
- Groups with large numbers of entries (e.g. Museums) should be able to bulk upload entries especially where there are many entries that have similar metadata. In addition, it may be necessary to make global changes to entries e.g. change to email address when contact has left. It should also be possible to bulk export metadata e.g. for editing or submission to other catalogues, etc.
- Search engine optimisation (SEO) will be necessary to enable the catalogue to appear in Google and other search results.
- Catalogue entries should be checked periodically for completeness and users should be notified of errors and omissions by email. This will be important as standards in sample archiving are developed to enable the catalogue to remain up to date.
- Feedback, metrics and user statistics may be required e.g. list of most updated and download statistics. Feedback should include contact us form (for collation of improvement suggestions), email links (to contributor, responsible party, point of contact) and list of contributors to expand the network. Monitoring and reporting may be a useful feature where users can collect operation information such as number of records added, deleted, viewed.
- A deletion policy may not be required since entries can revert to private status. Deletion of duplicate entries may be necessary.
- Questions were asked regarding whether entries should auto-publish and should there be incentives available to those that enter data and what would those incentives be?

The MuShCoW list on functionality that was compiled at the workshop is given in Appendix 6.4 (Section 6).

4.4 Web portal

The idea of a web portal/website that contained the discovery catalogue was discussed in plenary and was widely thought to be essential. The website could host items such as forums, blogs, etc.

Workshop attendees agreed that:

- A virtual UK-ESB would benefit from a web front end with other functionality.
- The website could feature entries e.g. recently updated/ added, most downloaded, etc.
- The website could include other functions such as clickable logos to other catalogues and groups as well as links to the websites of all contributors via a links page.
- The website could contain information on best practice, links to attachments and information on standard operating procedures.
- Some thought should be given to the name of the ESB. Some workshop delegates felt that specimen was not the correct term (although this is the term used by other “specimen” banks).

The MuShCoW list on the web portal that was compiled at the workshop is given in Appendix 6.5 (Section 6).

5 Conclusions and next steps

This project has further identified a need for, and developed the concept of a virtual UK-ESB. This primarily would consist of a web-based, searchable metadata catalogue to which specimen holders could readily enter their own metadata through the Web. The catalogue would be housed on a web-portal that would provide some additional general information and links to individual specimen holdings.

The key part of this project was to develop the design of the data-entry fields, search capability and general functionality of a virtual UK-ESB. This has been done in conjunction with key stakeholders through a workshop. While the requirements for the data entry fields and functionality were clearly articulated at the workshop, the details of the search functions were less well defined. Following the workshop an updated list of potential fields was developed that incorporated workshop feedback and communication with other interested parties (Appendix Table 6.6).

Development of the search functions should take into account the ease of use of search engines currently used on any other publically searchable ESBs from around the world. Overall, the process of consultation should help ensure “buy-in” from stakeholders and generally facilitate easy entry and searching of data.

The development of a UK-ESB is also likely to foster a national network or partnership of organisations that archive nationally important environmental specimens. It would also be expected that this partnership would extend internationally through the International Environmental Specimen Bank Group (IESB).

The next step for the development of a virtual UK-ESB is to implement the design and development ideas captured in the current report and build a test version. This would be trialled for ease of data entry and ability to search and involve stakeholders who have expressed ongoing interest in a UK-ESB. Once tested, the virtual UK-ESB would then be ready for launch and would need: (a) an accompanying website to be designed, (b) implementation of a communication strategy to disseminate information about its existence, and (c) an agreed maintenance programme. Support for such work could be support as part of the National Capability remit of NERC, or be met from a consortium of community funding, given the UK-ESB would meet and serve community need.

Given the close link to the UK-EOF (entries on the UK-ESB would be expected to come from a subset of contributors to the UK-EOF), it would seem logical to link the UK-EOF and a UK-ESB and to co-brand them. Maintenance, troubleshooting, development and review of the UK-ESB, and the participation in wider groups, could therefore potentially be done by the same support team that carry out the same functions for the UK-EOF. In this way, the UK-ESB would make a full contribution to the UK National Capability.

6 Appendices

Appendix 6.1. List of those that were willing to look at and comment on mock-up UK-ESB catalogue screens.

Name	Email address	Affiliation
Alex Tate	ajtate@bas.ac.uk	British Antarctic Survey
Alexandra Tomlinson	Alexandra.Tomlinson@fera.gsi.gov.uk	Fera
Alison Hester	Alison.Hester@hutton.ac.uk	James Hutton Institute
Amber Vater	amber.vater@ukeof.org.uk	UK-EOF
Andrew Johnson	ajo@ceh.ac.uk	CEH
Andrew Kitchener	A.Kitchener@nms.ac.uk	National Museums Scotland
Brian Etheridge	brian.etheridge@rspb.org.uk	RSPB
Chris Shortall	chris.shortall@rothamsted.ac.uk	Rothamsted Research
Colin Mackechnie	cmackechnie@ceh.ac.uk	CEH
David Cotton	dcott@oceannet.org	MEDIN
David Howard	dhoward@ceh.ac.uk	CEH
Dylan Lloyd	dy.lloyd@ccw.gov.uk	CCW
Elizabeth Chadwick	chadwickea@cf.ac.uk	Cardiff University Otter Project
Graham Rotheray	g.rotheray@nms.ac.uk	National Museums Scotland
Hardy Schwamm	hschwamm@fba.org.uk	Freshwater Biological Association
Heinz Rudel	heinz.ruedel@ime.fraunhofer.de	Fraunhofer IME
Helaina Black	Helaina.Black@hutton.ac.uk	James Hutton Institute
Helen Peat	hjpe@bas.ac.uk	British Antarctic Survey
Inma Robinson	inmbin@ceh.ac.uk	CEH
Jacky Chaplow	jgar@ceh.ac.uk	CEH
Jan Koschorreck	jan.koschorreck@uba.de	Federal Environment Agency
Jeremy Giles	jrag@bgs.ac.uk	BGS
Jerry Herman	j.herman@nms.ac.uk	National Museums Scotland
Jo Rae	jrae@bas.ac.uk	British Antarctic Survey
John Davy-Bowker	JDavy-Bowker@fba.org.uk	Freshwater Biological Association
Kevin Jones	k.c.jones@lancaster.ac.uk	Lancaster Environment Centre
Michael Taylor	Michael.Taylor@sasa.gsi.gov.uk	SASA
Mike Brown	mjbr@ceh.ac.uk	CEH
Mike Dobson	director@fba.org.uk	Freshwater Biological Association
Mike Howe	mhowe@bgs.ac.uk	British Geological Survey
Monika Jurgens	mdj@ceh.ac.uk	CEH
Paul Duff	p.duff@vla.defra.gsi.gov.uk	Veterinary Laboratories Agency:
Richard Harrington	richard.harrington@bbsrc.ac.uk	Rothamsted Research
Richard Shore	rfs@ceh.ac.uk	CEH
Rob Rose	rjr@ceh.ac.uk	CEH
Robbie McDonald	Robbie.McDonald@fera.gsi.gov.uk	FERA
Roy Nielson	Roy.Nielson@hutton.ac.uk	James Hutton Institute
Sabera Patel	sabera@ceh.ac.uk	CEH

Simon Turner	simon.turner@ucl.ac.uk	University College London
Steve Hughes	shug@ceh.ac.uk	CEH
Steve Ormerod	Ormerod@cardiff.ac.uk	Cardiff University
Susan Chambers	s.chambers@nms.ac.uk	National Museums Scotland
Thomas Maes	thomas.maes@cefasc.co.uk	CEFAS
Vicky Kindemba	vicky.kindemba@buglife.org.uk	Buglife
Zena Floody	z.floody@nms.ac.uk	National Museums Scotland

Appendix 6.2. Screen mock-ups and associated fields for virtual UK-ESB data entry screens

Screen 1. Sample description fields

Home -> Entry Editor -> Description Submit **1**

Description | Storage Information | Categorisation/Links | Contacts

Title
Help: title of sample collection

Sample Type (format)
Help: E.g. soil, plant, bird tissue, mammal tissue, plant, root

Sample Description (abstract)
Help: enter details about the sample collection.

Geographic Provenance (coverage-spatial extent)
Help: enter coordinates of a bounding box showing where samples were collected E.g UK, Cumbria, field site.



Timespan (coverage - temporal extent/reference date)
Help: Enter the start date of sampling regime or project and the end data when applicable.

start

end

Lineage (source)
Help: Information about where the samples came from, how they were collected, QC, QA, SOPs, etc.

N.B. The fields below are required to comply with standards but may not be visible to those creating a catalogue entry. They will autogenerate or default fill.

Identifier
Help: URI (a string of characters used to identify a resource on the Internet)

Language
Help: default - English (required to comply with standards)

Topic Category
Help: default - Environment (required to comply with standards)

1 Content will be submitted in a format which can be searched

Fields associated with Screen 1

<i>Field name</i>	<i>Alternative (comply with standards)</i>	<i>Help: Information on what is required</i>	<i>For example</i>
Title		Title of sample collection	Predatory Bird Monitoring Scheme Tissue Archive
Sample type	Format		soil, plant, bird tissue, mammalian tissue, plant, root, etc
Sample Description	Abstract	Enter details about the sample collection	
Geographic Provenance	Coverage - spatial extent	Enter coordinates of a bounding box showing where samples were collected	UK, Cumbria, Field site
Timespan	Coverage - temporal extent/reference date	Enter the start date of the sampling regime or project and the end date when applicable	
Lineage	Source	Information about where the samples came from, how they were collected, QC, QA, SOPs	
The next 3 fields are required for the catalogue to comply with standards (ISO19115 and Dublin Core) but may not be visible. They will auto generate or default fill.			
Identifier	URI	String of characters used to identify a resource on the internet	
Language		Default: English	
Topic category		Default: Environment	

Screen 2. *Storage Information* data loading screen

1 Content will be submitted in a format which can be searched

Fields associated with *Storage information*

Field name	Alternative field name (comply with standards)	Help: Information on what is required	For example
Sample availability	Rights	How to gain access	Email a@ceh.ac.uk for information
Access restrictions (O)		<i>IPR, T&C</i>	<i>Requests are dealt with by x or on a case by case basis</i>
Storage requirements/methods (O)		<i>Size, type of container, time taken to remove from store, specialist facilities required</i>	
Archive location (O)		<i>Location of archiving facility, freezer, research station</i>	
Sample history (O)		<i>Details of previous use</i>	
H & S issues (O)		<i>Information users need to know if they want access to samples</i>	<i>Stored in alcohol, under liquid nitrogen</i>

Screen 3. *Categorisation, keywords and links* data entry screen.

1 Content will be submitted in a format which can be searched

Fields associated with *Categorisation, keywords and links*

Field name	Alternative field name (comply with standards)	Help: Information on what is required	For example
Keywords	Subject	search a thesaurus for keywords	
Keywords	Subject	Add other unspecified keywords	
Resource locator	Citation	publications related to samples, archives	
Resource locator	Links	links to existing catalogues, websites	
<i>Additional notes (O)</i>		<i>other information about the collection, things that sample users might need to be aware of</i>	

Screen 4. *Contact information* page data entry screen.

- 1 These options only show up when the user is logged in 2 Contact details of the individual/organisation who entered the sample information 3 Contact details of the individual/organisation who are responsible for the sample information

Fields associated with *Contact information*

Field name	Alternative field name (comply with standards)	Help: Information on what is required	For example
Metadata point of contact	details about the person who is filling in the catalogue record		Email address, name, phone number and organisation name and address
Metadata date		Default: date the catalogue information was created	
Responsible party	contributor	Point of contact if different from above	contact this person for access to samples, archive facility, further information
One further field is required to comply with standards – auto fill/may not be visible			
Responsible party	publisher	point of contact if different from above	this person publishes the existence of the archive, sample collection

Appendix 6.3. MuShCoW for data entry screens and fields

Data entry fields
Must have
Unambiguous field names to avoid confusion
Dynamic help available both brief and in full
Boxes for data entry that are limited in size for different fields
Limit information on screens and number of tabs/screens
A defined level of granularity to users/contributors
Spell out access restrictions
Use different data entry types (buttons, drop downs, buttons, etc.)
Filter and sort category entries for the benefit of searches
Avoid duplication between tabs/screens
Should have
Offer guidance on the amount of information needed
Provide different forms of defining location region of samples
Provide indication of cost implications
Collect information to present from a users perspective
Use existing thesauri
Use autofill wherever possible
Credit funding bodies
Resource Type – to identify what level of description it is e.g. collection, item, a whole repository
Size of the collection e.g. Number of Items
Information & custodial history – how /when it got into the repository as opposed to initial collection information
Physical condition – effect on access or/and trigger for conservation work
Include preferred wording for citation for the collection
Include names of expeditions or projects
Accrual rate or not accruing
Date of original entry & who by but also date and who last updated it
Personal names of the collectors & the collecting organisations should be entered separately from the contacts list
Format for data entry should be controlled e.g. Surname, first forename second forename
Credit funding bodies
Include preferred wording for citation of the collection
Could have
Identify location of archive as a name or website
Reduce details of Health and Safety issues
Use buttons for opening text boxes for additional input

Appendix 6.4. MuShCoW for functionality

Functionality
Must have
Authentication. No requirement for a gatekeeper to add, edit and update information - enable user group to set up and maintain records. Record not necessarily owned by individual, ownership can be transferred.
Bulk upload, export and update – E.g. global changing all the email addresses for an organisation when contact has left. Ability to export metadata to other catalogues and tools.
Search engine optimisation (SEO) to enable Google, etc to find the website in its searches.
Validation on input
Periodic checks (e.g. auto generation of annual email)
Should have
Workflow i.e. save as draft, submit for publication, submitted, approved
Deletion for duplicate records although records can revert to private
Feedback <ul style="list-style-type: none"> • Contact us form to collate improvements • Email link to contributor (Responsible Party). • List of contributors to expand the network • Point of contact. • List of Most updated/Downloaded
Could have
Usage statistics
Rollback to previous version (future versions)
Links to other catalogues that link to us
Ability to tick a box to be notified later about records left in draft so that they can be removed or published.
Workflow Publication - telephone directory for content might be useful if errors present in the record.
Email notification to users to check validity of the record.
Monitoring and Reporting - product can collect operation information such as # of records added, deleted, viewed.
User feedback
Won't have
Deletion policy - not required since there will be workflow with publication status. Records can be reverted to private
Questions
Do entries auto publish?
Should there be carrots for entering catalogue data?

Appendix 6.5. MuShCoW for web front end

Web front end
Should have
Clickable links to logo
Page of links to websites of all contributors
Could Have
Best practice
Featured entries
Recently added/updated

Appendix 6.6. Updated list of fields (with explanation) that incorporates feedback from interested parties and attendees of 2012 workshop.

Field name	Explanation of field name and detail on what information is required
Reference numbers	Unique ID, UKEOF catalogue number, other catalogue number
Title	Title of sample collection
Sample type (format)	Information on what the sample/collection consists of.
Resource type (O)	<i>Level of description e.g. collection, item, whole repository.</i>
Size of Collection (O)	<i>i.e. Number of Items</i>
Sample Description (Abstract)	Detailed information about the sample collection. This may include physical condition (may affect access and trigger conservation work), expedition or project name and number, funding body, preferred wording for citation of the collection. Alternatively, this information could be added as keywords.
Collector name(s) (O)	<i>Controlled format e.g. Surname, first forename second forename – so that you can search for names consistently.</i>
Collecting organisation(s) (O)	<i>Controlled format e.g. name, address, postcode</i>
Geographic Provenance (Coverage - spatial extent)	Coordinates of a bounding box showing where samples were collected and/or place name (i.e. Town, County, Country) and/or site name.
Time span (Coverage - temporal extent/reference date)	Start date of the sampling regime or project and the end date where applicable
Acquisition information and custodial history (O)	<i>How and when it got into the repository as opposed to initial collection information</i>
Lineage (Source)	Detailed information on how, why and where the sample was collected, prepared, analysed, etc.

Identifier (URI)	String of characters used to identify a resource on the internet
Language (Required for standards)	Default: English
Topic category (Required for standards)	Default: Environment
Sample availability (Rights)	Is the sample available or already in use by another party?
<i>Access restrictions (O)</i>	<i>Intellectual property rights (IPR), Terms and Conditions, Environmental information Regulations (EIR) – may be required for exceptions; legal basis for restricting access.</i>
<i>Storage requirements/methods (O)</i>	<i>Size, type of container, time taken to remove from store, are specialist facilities required</i>
<i>Archive location (O)</i>	<i>Location of archiving facility, freezer, research station</i>
<i>Sample history (O)</i>	<i>Details of previous use</i>
<i>Health and Safety issues (O)</i>	<i>Information users need to know if they want access to samples e.g. stored under Liquid nitrogen, in alcohol, biohazard.</i>
Keywords (Subject)	Search a thesaurus for keywords
Keywords (Subject)	Add other unspecified keywords
Resource locator (Citation)	Publications that have been produced after use of samples, archives or collections, contain URI of catalogue entry or cite catalogue.
Resource locator (Links)	Links to existing catalogues, websites addresses, related collections.
<i>Additional notes (O)</i>	<i>Other information about the collection, detail that sample users might need to be aware of i.e. whether/how specimens are marked</i>
Metadata point of contact*	Details about the person who filled in the catalogue record.
Metadata date	Autofill with date the catalogue entry was created
Responsible party (contributor) - default* unless user fills in. (Required for standards)	Point of contact (anyone else who has contributed to record)
Responsible party (publisher) – default*unless user fills in (Required for standards)	Point of contact (person that makes the catalogue entry public)
<i>Accrual rate or not accruing (O)</i>	<i>Will the collection be added to?</i>
<i>Date of last update (O)</i>	<i>Date</i>
<i>Updated by (O)</i>	<i>Name of person who last updated entry</i>

Optional fields are italicised and denoted (O), fields that are required only to comply with standards are marked.