



ALERC
Association of Local Environmental
Records Centres

**A: ALERC Conference
Workshop Report**

A02

Data Flows Workshop

Paula Lightfoot, Mandy Rudd & Aisling Carrick

Association of Local Environmental Records Centres.
Conference 2011

1. Online Data Entry Portal for Ecological Consultants

Paula Lightfoot

2. Managing Data Flow from Online Recording Websites - Focus on Mitten Crab Recording

Aisling Carrick

3. Data Flow between Local Records Centres

Mandy Rudd



NBN 
National Biodiversity Network

Community Interest Company: Incorporation number 06951023 2nd July 2009
Registered Office: 13 St. Andrew's Crescent, Cardiff, CF10 3DB.
Website www.ALERC.org.uk. Forum: <http://forum.lrcs.org.uk>

Metadata

Addressee	ALERC Conference Directors: ALERC cic NBN Trust
Coverage.spatial	U.K.
Creators	Paula Lightfoot NBN Trust Data Access Officer Tel: 07930 458737 Email: p.lightfoot@nbn.org.uk Mandy Rudd GIGL mandy.rudd@gigl.org.uk Aisling Carrick Biological Data Officer Cofnod Intec, Ffordd y Parc, Parc Menai, Bangor, Gwynedd. LL57 4FG Tel: (01248) 672603 Aisling.Carrick@cofnod.org.uk
Date.issued	October 2011
Format	Adobe portable document (pdf)
Identifier	ALERC WG.A.02
Mandate	A. Memorandum and Articles of Association: ALERC cic. Incorporation number 06951023 13: Records to be kept, parts 1a, 1b, 2a, 2b, 2c
Publisher	Association of Local Environmental Records Centres cic National Biodiversity Network Trust
Rights.copyright	Association of Local Environmental Records Centres cic National Biodiversity Network Trust
Source	Association of Local Environmental Records Centres cic 2011 Conference
Status	Version b Proof
Subject	Report from ALERC Conference Workshop
Title	Data Flows 1. Online Data entry portal for Ecological Consultants 2. Managing Data Flow from Online Recording Websites: Focus on Mitten Crab Recording 3. Data Flow Between Local Records Centres 26th October 2011
Type	Conference Workshop Report
Digital Signature	

Data Flows Workshop: Case Study 1

Online Data Entry Portal for Ecological Consultants



Facilitator: Paula Lightfoot (NBN Trust)

Introduction:

It was recently estimated that between £100-200 million per year is spent on acquiring ecological data in the private sector¹. Professional guidance states that consultants should “*wherever possible, make scientific data collected during the course of their professional duties available to others such as records centres²*”, but most of these data are ‘lost’ in consultancy reports rather than being mobilised via LRCs to the Gateway.

It has been suggested by the consultancy sector that “*a common data entry gateway*” with “*automated reposting to the appropriate LRC^{2,1}*” would make it easier for consultants to share data.

Defra have agreed to fund the development of an online data entry portal for ecological consultants, with the aim of increasing data mobilisation by the commercial sector, particularly of protected species records.

It is hoped that ecological consultants will use an online data entry portal provided that it is straightforward and not time-consuming, especially if the portal processes their data into report-ready format for them. Use of the portal could perhaps be made a condition of their protected species licence.

For discussion:

1. Can you draw together on the flipchart pages one or more data flow models to show:
 - Where the portal should be hosted (This can be more than one website!)
 - Where the data should be stored
 - Who should be the data custodian responsible for onward data flow and access controls
 - How validation and verification should take place
 - How the data should be made available for use at local and national level
2. How could consultants be encouraged to use the portal, and who should be responsible for providing training and support?
3. How can Local Records Centres benefit from this project?
4. What risks or problems do you foresee with this project?
5. *What would be the best way to treat records of species which have been translocated from a site which has been developed? Are these records useful to LRCs?*
6. *What would be the best way to treat absence records? Are these records useful to LRCs?*

¹ Tasker, A. (2010) ‘Where is all the data going? An extra view’ *In Practice* Institute of Ecology and Environmental Management, Number 70

² IEEM (2007) Code of Professional Conduct: Guidance to Members. Institute of Ecology and Environmental Management, Professional Issues Series 5

Data Flows Workshop: Case Study 1

Online Data Entry Portal for Ecological Consultants

Participants

Charlie Barnes	Lincolnshire Biodiversity Partnership
Martin Horlock	Norfolk Biodiversity Information Services
Catherine Burton	Surrey Biodiversity Information Centre
Lois Brown	Warwickshire Biological Records Centre
Jez Elkin	Buckinghamshire and Milton Keynes Environmental Records Centre
Sue Timms	Leicestershire and Rutland Environmental Records Centre

Key points from discussion

1. A draft data flow model (fig. 1) was developed from the discussions held during the workshop.

It was agreed that there could be many different front-ends (portals) provided that these link to a common back-end (data 'warehouse').

The following websites were suggested as places where portals could be hosted:

- ALERC
- IEEM (but if a portal is hosted on IEEM's website, could it still be used by consultants who are not IEEM members?)
- Biological Records Centre
- NBN
- Individual LRCs
- ALGE's Planning Portal (Mike Oxford).

It was suggested that the data could be stored at the Biological Records Centre.

The following options for data custodianship were discussed:

- BRC could be the data custodian responsible for onward data flow and setting access controls but this would need to be done in agreement with ALERC and IEEM. This would probably be the most effective solution as data access could be set per survey rather than per taxonomic group or per geographic area.
- NSS could be the custodian responsible for data on particular taxonomic groups, but only if this does not restrict LRC access to the data.
- LRCs could be data custodians for the records within their geographic area, but a survey dataset might overlap neighbouring LRC boundaries.

It was queried whether data from consultants actually need to be verified at all, as they are license-holding professionals. Validation could be largely automated and verification could be carried out online by LRCs and NSS experts logging into the system.

The data could be made available for national use via the NBN Gateway. They could be made available for local use via LRCs, ideally using Web Services to stream the data from the NBN Gateway rather than holding copy datasets.

2. The following suggestions were made for encouraging consultants to use the portal:
 - Discounts on data searches from LRCs for consultants who submit data via the portal. This might not be sustainable at a time when public sector funding is being cut.
 - It could be made a condition of protected species licenses that the license holder submit data via the portal. Many types of survey do not require a license so this would not be sufficient in itself.

Data Flows Workshop: Case Study 1

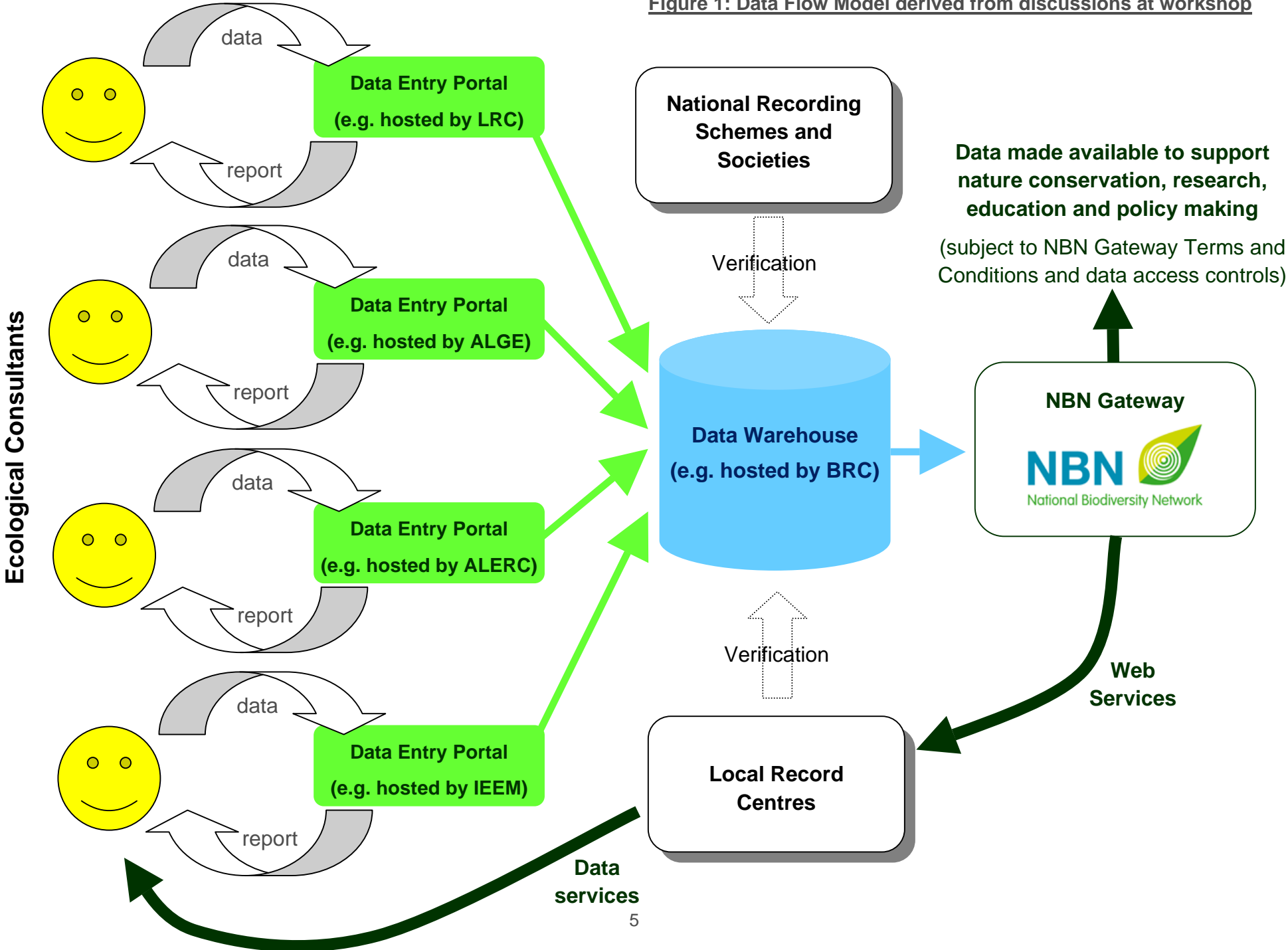
Online Data Entry Portal for Ecological Consultants

- The availability of apps or custom-designed portable electronic devices for ecological surveys, enabling a consultant to enter records in the field and download them in the office in report-ready format. Date, time, position and recorder's name would be captured automatically. Species names would be selected from the Species Dictionary to avoid ambiguity and capture designation information.
 - The portal should be able to accept both manual entry of individual records and upload of a complete list of records, e.g. in Excel, and collate these into report-ready format.
 - IEEM could encourage members to share data via the portal by setting targets for data submission in the same way that they set targets for CPD. The portal could provide a summary to each consultant of how much data they had submitted and IEEM could collect information annually in the same way that they collect information on each member's CPD.
3. It was felt that Local Records Centres could benefit from the project by gaining more efficient access to consultancy data, especially on protected species – currently these records are often extracted manually from survey reports provided to the Local Planning Authority which is a time-consuming process.
4. The following risks were identified with this project:
- It is important to consider the full range of ecological consultants' requirements for formatting data into survey reports, e.g. consider marine and freshwater as well as terrestrial survey methodologies
 - An increase in data mobilisation from the consultancy sector could place a burden on those able to verify data, e.g. VC Recorders, and this could create a bottleneck. It was acknowledged that Record Cleaner might help to reduce this burden.
 - Duplication could arise if LRCs manually extract species records from survey reports obtained through the planning system.
 - If consultants start sharing more data they might resent paying LRCs for data searches, even though it is not the data they are paying for but the LRC's services in data management and supporting local recording.
 - Consultants must have full access to their own records via the Gateway, but not to records shared by other consultancies. This would be possible under the proposed new system of Gateway data access controls but is not possible at the moment.
 - LRCs and Country Agencies must have full access to the data via the Gateway to ensure the records are used to support planning, land management and policy making at a local and national level.
 - The public must not have full access to the data via the Gateway as this could undermine LRC business models and client confidentiality.
 - Long term funding is vital to manage and maintain the online recording infrastructure developed under this project.
 - If records are made available to LRCs only via the Gateway, capture resolution must not be lost. The Gateway currently displays records at a maximum resolution of 100m (6 digit grid reference), but a bat roost or veteran tree might be recorded to 10m or even 1m resolution.
 - The parameters of the project need to be clear – will this be limited to species records or is there scope to develop it to gather habitat data in future?

Data Flows Workshop: Case Study 1
Online Data Entry Portal for Ecological Consultants

- The data need to be made available promptly via the NBN Gateway. Delays could arise if consultants wait until the survey season is over before sharing data or if the records cannot be verified quickly.
- If data sharing is not 'enforced' (e.g. by IEEM or the Country Agencies) then it is absolutely vital that there are genuine incentives for consultants and that these are well publicised, because consultants do not cost data sharing into their budget.
- It would be beneficial if LRCs could access data from the NBN Gateway via Web Services, but very few LRCs are currently using NBN Web Services so this would require investment of time and financial resources.

Figure 1: Data Flow Model derived from discussions at workshop



Data Flows Workshop: Case Study 2
**Managing Data Flow from Online Recording Websites:
Focus on Mitten Crab Recording**

Facilitator: Aisling Carrick (Cofnod, ALERC)

Introduction:

There has been a recent increase in the number of online recording websites, and this situation is only likely to increase over time, one reason being their active promotion by at least one of the country agencies (Natural England). These systems are very varied, and range from a small number set up and managed by Local Records Centres (LRCs) (e.g. rECOrd, Cofnod), to which recorders can add records of any species, to national ones focussed on particular species groups or even single species (e.g. BTO's BirdTrack, the Recording Invasive Species Counts (RISC) project). This online data capture offers an alternative to the traditional methods employed by LRCs, National Schemes and Societies (NSS) and other bodies, and can provide recorders with a number of useful tools including online mapping facilities and access to the NBN Species Dictionary. The ability for records to be rapidly viewed and verified by relevant experts is often built into these systems.

The proportion of records flowing through these online recording websites will inevitably increase over time. What does not appear to have not received much consideration so far, however, is how this is likely to affect data flow between LRCs, NSS, the NBN Gateway and other relevant organisations.

Focus on Mitten Crab Recording:

One recent example is the Mitten Crab Recording Project website (set up by a partnership including the Natural History Museum, Countryside Council for Wales (CCW) and The Marine Biological Association (MBA)), which was launched last week with an online recording facility (actually part of the RISC online recording facility). One of the key objectives of this initiative is to ensure that records of Mitten Crab come to marine specialists, e.g. within CCW, as quickly as possible. The website is available at <http://mittencrabs.org.uk/>. Data flow mechanisms between Welsh LRCs and this project are currently being discussed, but this could be relevant across the UK and also as a case study for other similar scenarios, especially where speed of data transfer is crucial.

The key participants/nodes in this data flow model are: Individual recorders/members of the public, Individual LRCs, the Mitten Crab/RISC project, the Marine Biological Association, the NBN Gateway and the country agencies. The aim is for the MBA to hold the 'top/master copy' of any Mitten Crab records, within the DASSH (Data Archive for Seabed Species and Habitats) database which they manage. The data feed between the Mitten Crab/RISC website and DASSH is instant, so they can be regarded for the purposes of this discussion as one node in the data flow model. It is possible that data may flow best through several different routes depending on the circumstances of a particular LRC/Country Agency for example.

Data Flows Workshop: Case Study 2
**Managing Data Flow from Online Recording Websites:
Focus on Mitten Crab Recording**

For discussion:

1. Does your LRC collect records of Mitten Crab (or other invasive species)?
2. If so, are they submitted to the MBA/DASSH and/or to the NBN Gateway, and how regularly?
3. Would you consider submitting them to the Mitten Crab/RISC website?
4. Can you see any potential issues with transferring custodianship of records to the MBA/DASSH?
5. Does your LRC supply records of Mitten Crab (or other invasive species) in response to data enquiries?
6. If so, how important is it that you receive new records rapidly?
7. How would you prefer to access mitten crab records in your area, e.g. for use in client data searches:
 - a. Holding a copy of the record in your database (flagged as uncertain until it has been verified by the MBA)
 - b. Receiving verified data via web services from the MBA
 - c. Receiving verified data via web services from the NBN Gateway
 - d. Other suggestions
8. Can you draw together (on one of the flipchart pages) one or more proposed data flow routes, based on the discussions above?

Key points from discussion

Not all participants were aware of any Mitten Crab records having been collected by their LRC (as yet). The group's discussions centred around the design of an appropriate data flow model (see next page), including the need for data to flow rapidly to all the key participants. Two data flow routes have been shown: data would be likely to flow using a combination of these.

Data Flows Workshop: Case Study 2
**Managing Data Flow from Online Recording Websites:
 Focus on Mitten Crab Recording**

Suggested Data Flow Model

1: Ideal data flow →

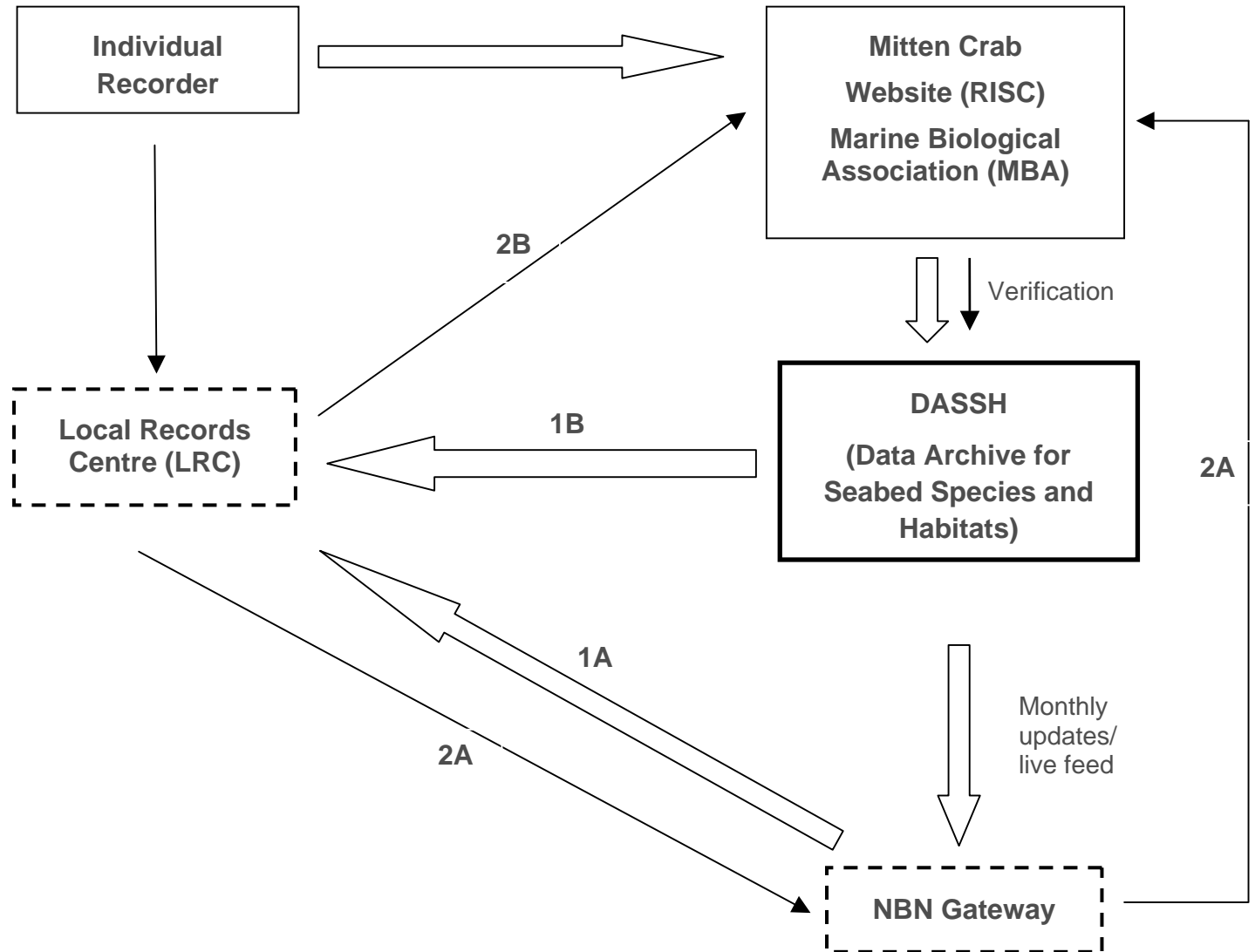
An individual recorder approaching an LRC with a Mitten Crab record is asked to submit it via the Mitten Crab website (or the LRC submits the record on their behalf). The top copy record is held by DASSH.

The LRC accesses the record via the NBN Gateway (1A, data download or web services) or more rapidly via direct 'download feed' from DASSH (1B).

2: Alternative data flow →

Individual recorders submit records directly to an LRC (which may happen to some extent in mixed datasets anyway)

MBA accesses a copy of the records via the NBN (2A), or by direct data exchange with the LRC (2B).



Data Flows Workshop: Case Study 3

Data Flow Between Local Records Centres

Facilitator: Mandy Rudd (GIGL, ALERC)

Introduction

The London and South-East LRC Group has set out its position on the operational structure of the individual centres within the 2 regions. It is split into 2 main functions:

Data collation – each LRC will work with organisations and individuals who generate data for its operational area (each modern county / regional boundary). If an LRC's data providers are also recording in a neighbouring county, they are offered the opportunity of providing data to the original LRC with the agreement that their data will be exchanged at the LRC level. This ensures each LRC has access to data for their area, and the data providers don't have to collaborate with more than one LRC unless they wish to.

Data provision – the LRC's modern county remit sets out the area in which it conducts business / seeks funding partners. Any data searches that cross county boundaries are undertaken and neighbouring LRCs' services flagged in order that the customer gains access to all data for their search area.

For discussion

1. Could recognition of geographic data flows strengthen LRC partnerships and prevent duplication at a UK level e.g. on the NBN Gateway? E.g. if EA, NE, CCW, SNH etc actively support their LRC partnerships in focusing local data flow would it make the LRC community stronger/more efficient within the NBN partnership?
2. What technological solutions are there to exchange data rapidly between LRCs?
3. How do we avoid duplication at LRC level and on the NBN Gateway?
4. How does this fit in with LRC accreditation?
5. What are the implications for sharing data across country boundaries? E.g Cofnod and rECOrd
6. What local, regional and national agendas is this topic relevant to?

Participants

Teresa – Cumbria

Tom – Cheshire

Steve – Greater Manchester

Alexa – Fife

Katherine – Northumberland

Lisa - Dumfries

Key points from discussion

1. Solutions for cross-boundary data searches
 - a. LRCs to share data within a buffer area based on administrative boundary e.g. Greater London to provide Greater London data to Surrey and vice versa, with

Data Flows Workshop: Case Study 3

Data Flow Between Local Records Centres

the agreement that the central grid reference of the data search area determines who runs the search. So, if it falls in Greater London GiGL would undertake the search for the customer. This would streamline the process for consultants but would mean a small loss of income to Surrey (in this example).

- b. Or, both LRCs would do the data search for their part of the search area but at a reduced hourly rate in order to show willing to our customers
2. Technological solutions to sharing data across administrative boundaries include:
 - a. Bespoke online portals
 - b. Sending GIS files via email etc
 - c. Web services if they offer the option to tailor availability based on geography
3. Recommend 1:1 vice county and LRC relationships. E.g. a VC recorder only has to establish and maintain a relationship with one LRC for the purpose of support and data, and the LRC will share data for outside their geographic remit with neighbouring / other LRCs on the recorder's behalf

Further Qs or discussions:

1. Should ALERC have a UK-wide position on data sharing between LRCs for the purpose of clarifying our operations to common customers and partners
2. General point: could we start utilising the regional LRC meetings to further ALERC ideas and development throughout the year? E.g. writing standardised progress reports per LRC prior to regional meetings that are then posted on the ALERC forum. Or producing ALERC documents on behalf of all members etc.
3. Could focusing an LRC data flow model on geography rather than taxonomy work? I.e. the NSS working in London would concentrate on taxonomic flow from their local reps in London to their national database* and NBN Gateway, whilst GiGL would focus on working with local / regional data producers and make their data available at a local and national level on their behalf. If the local organisations / individuals could be persuaded not to send their data to national organisations as well, it would eliminate the possibility of duplicates. This could potentially strengthen the LRC networks role both within their own administrative areas and within the National Biodiversity Network partnership as we would be recognised as the local facilitators of local data flow

*The LRCs could undertake data mobilisation for NSS that don't have the resources to mobilise data from local to national e.g. VC recorders send data to LRC, then the LRC passes data to the Gateway for use by the NSS.

Data Flows Workshop: Case Study 3 Data Flow Between Local Records Centres

Figure 1:- South-East England (inc. Greater London) showing current LRC operational structure and historic recording units

