

National Biodiversity Network Conference 2017: Workshop 5: What are the barriers to sharing data on the NBN Atlas and how can we overcome them?

Workshop aim

To explore what people and organisations view as the barriers to sharing data via the NBN Atlas, are these real or perceived and what can be done to alleviate concerns.

The workshop

Workshop attendees were separated into groups and asked to focus on what they felt was preventing data sharing to the NBN Atlas for Recording Schemes, Local Environmental Records Centres and the rest of the Network and what could be done to overcome these issues.

This report is a summary of group flipchart notes and from notes taken during feedback and discussion during the workshop. A personal summary has been provided at the end of the workshop notes.

A breakdown of the workshop aim were given as

- To understand what is preventing data sharing?
- Are the issues real or perceived?
- Can we give any case studies to satisfy concerns?
- To understand if the licences are trusted and if not, why not?
- To show that more data are being shared through the NBN Atlas than the NBN Gateway and why this is a good thing.
- Has data licensing helped with lowering the barriers?
- Do people really understand what is involved in sharing data through the NBN Atlas?
- Do people really understand the NBN Atlas?
- Do people accept the positive benefits of sharing data?

What are the barriers to sharing data on the NBN Atlas?

This question prompted the majority of the group discussion and was expanded to include the linked sub-topics around perception and some offers of solutions.

Key points raised included;

- 1. Permission to release:** A recurring concern raised from the groups, particularly in respect of full resolution and 'open licence';
 - Provided under a different (e.g. non-commercial use or non-standard) licence.
 - Historic data with legacy permissions and now deceased recorders/owners.
 - Data owners have consciously decided that their data should not be made open
 - The large amount of time and activity contacting data owners and seeking new permission

Suggestions

- Concise guidance and examples on how to approach and gain access from recorders would be useful
- Provide examples of the benefits of sharing biodiversity data to help convince recorders to allow their data to be 'open'
- Provide positive feedback and detail use of records which recorders have allowed to be open

- Recognise and resource the effort required gaining and encouraging permission to be granted.

2. Concerns around misuse of data: An expansive discussion point where concerns around risks to data supplier funding models were raised including both LERC and NSS data and interpretation services. The potential for misuse does not just come from commercial users but also exist around the accuracy of interpretation of data or deliberate misrepresentation as well as concerns from recorders around providing the precise location of species which may not be considered 'sensitive' but are none the less at risk from collection or persecution. Similar issues were also raised in the 2016 Workshop 4 'Improving use of our data'¹.

Suggestions

- Sustainable method of funding the support of local recorders and mobilisation/quality assurance of data at the local level, a grass roots approach. Could be delivered by a planning levy and so funded directly by those profiting from the Network. This could initially be considered as an 'opt in' option for planning authorities to assist in the resourcing of their statutory duties (needs based) with a remit for transparent and targeted use that meets the needs of the Network as a whole.
- Policing/Enforcement of the Creative Commons licences. It was noted that the Secretariat themselves would not have the resources to police data misuse. However, there are in place protocols to deal with breaches of the licences once they have been flagged up. A defined method for reporting and evidencing misuse is suggested so that the Network can take ownership of its own data either via the NBN Atlas or through Network professional partnerships.
- *Access controls.* Asking a general question to the floor on 'would attendees like controls reinstated' the clear majority response was yes. It should be stressed that there are complex reasons behind this response. Some organisations are unable to release information fully by request of the record owner, others have concerns as described above around misinterpretation and misrepresentation. However, a 'stepped access system' was clearly regarded as a solution.

3. Concerns over data quality: Data partners and Network organisations want to see the Atlas displaying good, fit for purpose, data and for those data to be used appropriately but concerns were raised about the processes data has to go through before it is accepted and presented. Who should the Atlas accept data from? How can we ensure that the data are properly verified and validated?

Suggestions

- Clearer data flow pathway.
- Clearer set of responsibilities for data providers.
- A focus on validation activity
- Clarity around verification

4. Confusion over responsibility: Discussion within the groups highlighted some confusion over aspects of data flow to the Atlas and who was responsible for sharing data, this was a

particular point of concern for individual recorders who are receiving multiple requests to record and share their data but are unsure of the best method of doing so, inevitably leading to duplication as records are shared to multiple organisations or the same one through multiple channels.

Suggestions

- Pathway and responsibility: Clarity around the data flow pathway, who to send records to and where they go, and responsibilities at each stage.
- Network collaboration: Reduce duplication of effort by collaboratively developing recording activity that benefits multiple Network members.
- Network consultation: Consult the Network on responsibilities for data collection, validation, verification and sharing to the Atlas.
- Inform recorders: Be clearer with 'recorders' at point-of-collection (online form notification, part of volunteer training/induction) as to where, how and at what resolution their records will be shared.

5. Other concerns: A variety of other concerns were also raised and noted through the feedback and notes;

- Time required to ensure data can be shared fully and openly via the Atlas
- Technical difficulties;
 - Conversion into an Atlas compatible format (especially for small recording schemes)
- Concerns around sharing of data that may have been collected illegally (e.g. trespass)

Licences – Do you trust them and are they working for you?

During the second part of the workshop groups were asked to focus on licensing, how they felt the system has helped and whether or not they fully understood it.

There was general feeling that licences were not trusted by the Network. However, it was recognised by most that a standardised licensing is more efficient and Creative Commons better recognised. It was also noted that machine readable licences have enabled data exchange with other systems such as the Global Biodiversity Information Facility (GBIF).

It was felt that there could be more visible and accessible statements as well as examples of the licences and what they mean to help data providers and users better understand why we apply them and how they should be used

Group feedback and flipchart notes are summarised below.

Concerns

- Don't know what they are
 - Not fully understood by recorders and organisations
- Not trusted

Suggestions and points raised

- CC licencing is required to improve confidence and efficiency in use of data
 - Administration of other non-machine readable options costly.
- Examples of use
- Short clear statements and examples/case studies
- Clarity and documentation more visible
- Dispensation for LERCs to enable them to use records licenced as CC-BY-NC
- Policing takes up time and resource, not effective option
 - Default settings by user type
- CC0/Open Licence
 - Enable focus instead on interpretive products/services/recorder support

Personal Summary

Barriers: A range of barriers were raised by the groups many of which have been explored in previous reviews and workshops¹. A number of the points raised have already been tackled to a greater or lesser extent (e.g. the sharing historic/legacy records², Biological Records and Intellectual Property Rights³) as has licensing (see below). Also, recent work by the Sussex Biological Records Centre (SxBRC) 'Sharing Sussex Beetle Records Lessons Learned'⁴, raised during the workshop, provides a factual account of recorder responses to the release of data openly on the Atlas and the process undertaken to release data and documents many of the concerns and issues raised during this workshop. Such studies could feed back to the Network to help advise a way forward.

My take from the discussion was that a lot of concerns came from the potential for misuse of data which would undermine the stability of the Network and its aims. Trust appears to be a constraint which largely goes un-mentioned but appears to be behind many of the concerns.

That said, 'resourcing' is a real issue and is putting increasing demands on an already stretched volunteer network and expecting organisations to deliver work using staff time that is not paid for.

Licensing: There was confusion over licensing and why the new licence system had been put into use. It's worth noting the Secretariat has published concise guidance on licensing⁵ but this could perhaps be better communicated to and by Network members, in particular to members of the Network that are not direct data providers.

Case studies, examples and trust: Network organisations are being asked to hand over into the public domain information in which they have heavily invested be it personally or organisationally. This requires a substantial amount of trust and cross sector collaboration.

There is a willingness to engage and a broad acknowledgement that acceptance that a national hub is a good means of achieving benefits for both Network members and their aims. However, this must be done collaboratively and with feedback, review and consultation to engender trust in the process.

We could do better by identifying examples of good and positive use of data not just of where data has enhanced its use but how that has fed back and added value to the work of the data providers.

The use of data object identifiers by GBIF could be an excellent vehicle for delivering this if it were to be made available by the Atlas and recognised as an industry requirement not limited to the academic sector (e.g. source your data!). Such transparency would also help to enforce the licences by enhancing accountability.

¹NBN Conference: Workshop 4 Increasing use of our data <https://nbn.org.uk/wp-content/uploads/2016/08/Workshop-4-Increasing-use-of-our-data-summary.pdf>

²Guidance for adding historic data <https://nbnatlas.org/help/guidance-adding-historic-data/>

³Further explanation of Intellectual Property Rights and how they pertain to biological data records <https://nbn.org.uk/wp-content/uploads/2017/03/Guidance-for-transfer-of-historic-data-Further-explanation-of-Intellectual-Property-Rights.pdf>

⁴Sharing Sussex Beetle Records Lessons Learned http://sxbrc.org.uk/documents/LessonsLearned_SharingSussexBeetleRecords.pdf

⁵Licences on the NBN Atlas <https://nbnatlas.org/help/data-licenses/>

⁶What are open licences? <https://theodi.org/guides/publishers-guide-open-data-licensing>

Group Flipchart Notes

Constraints

- Confusion over who is sending records
- Validation is undervalued, focus is all on verification
- Lack of funding is restricting LERCs ability to share data (confidence and business model)
- Permission may not have been given to share data – sometimes impossible to go back and ask all recorders. Historic data especially.
- Data protection legislation. Hard to find contacts for recorders
- Lack of trust
- Too many places to send records is confusing for recorders
- Lack of trust
- Illegal collecting (NE permissions?)
- Academic institutions want publications from dataset

Solutions

- Funding for LERCs
- Clarity to recorders around where their data will be sent and at what resolution
- Encompass 'access controls' into the Atlas so e.g. LERCs, Universities, Recording Schemes... can use discretion when providing their data. Mutual trust.

Open Data

- Case studies of where amazing things were made possible
- Case studies exploring the costs of locking down data (GBIF)

Misuse of data (not just commercial)

- Transparent example of action
- Not policed proactively by 'NBN Sec' need reassurances that procedure exists

Academic Papers – Using Data

- Acknowledgement of how data is used
- Automated thank you

Non-commercial Licence

- 'Pop-up alert'
- Filters automatically assigned based on user type

Feedback Notes

What is preventing full/open data sharing & how do we solve it?

- **Validation.** Organisations want good quality data on the Atlas. Data that can be trusted and a Network which they can be proud to be a part of. Currently Validation standards are seen to be lacking and/or undervalued
 - Solutions in the form of better joint working, RC update, more clarity and perhaps better standardisation. What needs doing, whose responsibility etc...
- Recording methods could be improved and software issues resolved
- Business models
 - Valuing aspects other than pure numbers of data
 - Planning Levy
- Access controls
 - Vote to re-instate
- Licensing
 - Not fully understood by recorders and organisations
 - Examples of use
 - Clarity and documentation more visible
- Misuse stress this does not just apply to LERC's and commercial misuse
 - Better awareness or flagging on Atlas and through professional and statutory partnerships
 - Some aspect of policing / tracking of use
 - Enforcement
- Is my data used?
 - Auto thank you / ability to trace usage
- Case Studies

- Positive stories around open data use and what can be achieved. Added value and return (?)
- Historic data and permissions/IPR
 - Legal explanations for the layman.

Licences – Do you trust them and are they working for you?

- Don't know what they are
- Dispensation for LERCs
- Short clear statements and examples/case studies
- People want data to be used but recognise that it needs resourcing
- CC0
 - Focus instead on products/services/recorder support
- Policing takes up time and resource, not an effective option
 - Not trusted
 - Default settings by user type
- CC licensing is required to improve confidence and efficiency in use of data (machine reading to improve data sharing and compatibility)
 - Administration of other non-machine readable options costly and ineffective.
- Licensing
 - Not fully understood by recorders and organisations
 - Examples of use
 - Clarity and documentation more visible