Treating Data Deficit Disorder - where next for the NBN?

By John Sawyer and Rachel Stroud
E: j.sawyer@nbn.org.uk

Presented at a workshop organised by the Macroecology Specialist Group of the British Ecological Society, held at Charles Darwin House, London on Monday 22 Sept 2014.

Introduction

Thank you for the invitation to talk here today. Rachel Stroud and I have prepared this talk about data deficit disorder and what that means for the future strategic direction of the NBN. Our thanks go to Paula Lightfoot of NFBR who provided some slides and Graham French of NBN Technical team who prepared many of the statistics for inclusion in the talk.

My interest in the natural world came from holidays spent at my family home on the isle of Mull on the west coast of Scotland. It is there I gained a love for observing, learning about and recording wildlife and where our family have been recording wildlife using the old style app called a notebook. And, along with a number of colleagues, I have built a portal about the biodiversity of the island called Wild Mull (www.wildmull.com). This has fact sheets for more than 1500 taxa and close to 3000 images.

The reason I mention my life on Mull and the work we have been doing on the Wild Mull website is because it could be considered a microcosm of the UK data deficits.By that I mean there are few records:

- Of species that few people know (i.e., fungi, insects and macro-algae)
- Above 300m and below sea level
- Common species are often overlooked (people like recording the rarities)

There are a few groups of species where we have mistaken ID and unidentified species (e.g. Euphrasia). We have found lots of dispersed data (unavailable via NBN due to access constraints and resolution issues). Finally, people are collecting data into their own systems (Excel, notebooks etc).

Data Deficit Disorder

The concept of Nature Deficit Disorder has been around for many years, coined by Richard Louv who wrote Last Child in the Woods. This is all about how young people are spending less and less time amongst nature and this is causing health and behavioural problems, as well as creating an ecologically illiterate population. The problem is so bad that some children don't know the difference between species or which are native or exotic. That leads us to Data Deficit Disorder which we are defining as:

Humans have limited access to data and information about the natural world and this has consequences:

Bad decision-making causing destruction / decline of significant ecological sites
Reduced opportunities to learn about ecology and ecosystem function
Reduced research capability and application impact
Inefficient natural resource management

This is our definition so please feel free to tell us how it can be improved. We should care about Data Deficit Disorder because it affects:

- **State of environment reporting**
- **Conservation / risk assessments**
- **Offsetting**
- **Nature's services**
- **Natural capital programmes**
- **Learning about wildlife**
- **Recording of wildlife**
- **Research programmes**
- **Wildlife conservation**

UK Data Deficits

We have put together a list here of 10 data deficits that we have observed. There are others (such as analysis deficit and interoperability deficit) but we thought we should start with just 10. We do not apologise for the first part of this talk sounding a bit negative. We hope by the end you will realise that there are plenty of solutions to most, if not all, of these. The deficits are:

- Geographic
- Taxonomic
- Temporal (time to market)
- Verification deficit
- Access and resolution
- Data loss
- Disaggregation
- BNFN
- Social (limited teamwork)
- Devolution Deficit

These combine to create an innovation deficit.

Geographic deficit

So, firstly 'Geographic deficit'. The 10km terrestrial grid square with the most records on the NBN Gateway is SK49 is Rotherham with 861,737 observations. But NB33 has only 1207 observations making it one of the least recorded grid squares in the UK. 367 grid squares have fewer records than NB33 but they are mostly over water. What that means is the hills behind Stornaway in the Outer Hebrides would be a great place to go for a field trip.

Other geographic deficits are the apparent absence (or non-recording) of very common species from many grid squares across the UK. Species such as rabbit and blackbird have geographic gaps in their distribution. Blackbirds are the most commonly recorded bird species on the Gateway and yet there are many gaps in the data over the last 5 years.

Recording absence is another Geographic Deficit issue whereby people do not always record what they don't see even though the NBN Gateway can store absence data.

Taxonomic deficit
There is also a taxonomic deficit which refers to the 78,179 taxa on the NBN Gateway.

Of those 29,376 taxa have no data (37.6%)

Includes extinct taxa not in UK such as Lynx lynx

Includes non-native taxa that for some reason are on the taxon list (e.g., Gorilla gorilla)!

13,482 taxa have 5 or fewer observations (17.2%) which means that 54.8% of all species on the Gateway have 5 or fewer observations.

6391 taxa have more than 1000 observations (8.2%) so our knowledge of UK biodiversity is really restricted to only 8.2% of the biota.

There is also another major gap in our habitat or ecosystem data holdings. Habitat data on the NBN Gateway is seriously in deficit. We know there is a diverse array of habitat inventory data types but much of the data associated with this is hidden or not digitised.

Temporal deficit
There is also a temporal deficit. That is the Time to Market for data from the day you record it to the time it travels through the system so that it can be consumed. How do we track a record through the system? That is not so simple and in March 2014 I loaded records to several systems but that data remains unusable because it has not arrived in the market.

A good example of this delay is illustrated by the NBN Gateway observation data for invasive non-native species. Those species are being recorded but the data are taking a year or so to make it across to the NBN. This affects conservation, environment decision making, research and peoples learning about wildlife.

Verification deficit
Verification deficits also occur although many people are doing an amazing job. For iRecord the mode verification time is 8 days although some records are not verified after 1 year. iSpot achieves some form of verification usually within an hour, with its advanced form of reputation-based validation. The NBN Record Cleaner system is working well in many cases but not being updated and not currently "owned" which is an issue we must address. We are finding that the verification team, which is largely made up of volunteers, is often overloaded.

Access deficit
There is also a serious Access Deficit on the NBN Gateway. Although we have 101 million records, 36 million of those records (from 87 datasets) are invisible (i.e., no public access). This includes protected species (which need to be protected), research capital of organisations that want to use the data before others do, and non-native species which is being blocked. Again using invasive non-native species on the Gateway as an example - 21% of all data is invisible. That is a significant problem if we are serious about tackling invasive species in the UK.

Data loss deficit
Data loss deficit is illustrated by research published in 2013 by Vines et al in Current Biology. They showed that 80% of research data are unavailable after 20 years and availability drops by 17%

each year after a paper is published. That means data is disappearing as we speak and that does not include undergraduate research and data captured by ecological consultancies.

Resolution deficit
There is also a resolution deficit in that of the 101 million records on NBN Gateway, only 12 million are publicly available at highest or full capture resolution. That means there is an 89 million observation deficit and that is just for the data that is on the Gateway.

Disaggregation deficit
There is also what we are calling disaggregation deficit. This is the breaking apart of the recording and verification community as a result of developing multiple entry points for the system. We have multiple recording systems (iSpot, iNaturalist, Living Record, Rodis, iRecord, LRC infrastructure). We also have multiple smart phone app systems (for ladybirds, Bird Track, Mammal Track, Wildlife Record, Plant Tracker, Rinse, Invasives etc).

We are breaking apart the recording and verification community in a way which is unsustainable and not very customer focussed.

By Nerds, For Nerds Deficit
There is also the BNFN deficit which means By Nerds for Nerds. To the un-initiated the landing page on the NBN Gateway when you click on 'Browse Species' is not ideal for the average user. Surely we can do better than that for the user, the data provider that is trusting us with their data and for the biodiversity that needs help.

Social deficit
There is a significant Social deficit and by that we mean a lack of team work. All these things have been said to me since I started in May.

Why should we put our data in the NBN?
What are you (NBN) going to do for us?
We are fine on our own – we don't need you
This is my data and I will take it to my grave
We make money out of that data so you cannot have it
People will steal our data if we give it to you
They are collecting there data all wrong – it's rubbish but we will not help them

If we really want to tackle the deficits we have to work as a team.

Devolution deficit
There is also a devolution deficit as UK fragments. This may not seem a problem after last Thursday's referendum vote but the UK has complex data gathering and reporting options. Countries have different biodiversity indicators, different invasive species priorities, different state of environment reporting.

NBN tackling the deficits
So – what does that means for the NBN? We are assuming here that we want to share data because we believe that it is in the best interest of the natural world. My colleague Alan Saunders who I worked with at the NZ Department of Conservation and in international consultancy work on the Juan Fernandez Islands said: Always ask: "*How will this benefit nature*?"

We think the NBN has a big role to play in fixing the deficits. The mission of the NBN is that: *We are an ever-growing network of individuals and organisations, recording and caring for the UK's wildlife data and making it universally accessible to the public, educators, researchers, conservationists and environmental decision-makers.*

When we talk about the NBN we am really meaning the wider group of organisations that contribute to the bigger vision and strategy. This is not just us as the Trust. As we are sure you know the NBN runs the NBN Gateway with 101 million records from 827 species datasets. We are a hub for more than 150 wildlife organisations - a family of people and projects with more than 100 NBN members.

The NBN is currently refreshing its strategy including its strategic aims. This means the NBN is focussing its aims on:

1. Capturing, diversifying and enhancing wildlife data
2. Making wildlife information available to those who need it
3. Captivating and engaging people about wildlife
4. Providing the best wildlife information management
5. Supporting the Network to support wildlife

In terms of tackling the deficits we are presenting some ideas here but we look forward to hearing your views as we renew the strategy so any input is welcome.

Deficit	*NBN action*
Geographic	*Data gaps promoted and plugged - but which are priorities?*
Taxonomic	*Taxonomic gaps promoted and plugged - - but which gaps should we target?*
Temporal	*Time and motion assessment to find out how data travels through our network, and work to clear blockages*
Verification	*Develop and support a Verification Network to ensure these people have what they need to do their job*
Access and resolution	*Capture more open data from people not currently involved and negotiate with others to open up data more in terms of access and resolution*
Data loss	*Work with science community and consultants to publish data. Undertake data maturity assessments to ensure people know they are curating their data following best practice*
Disaggregation	*Set standards (for attributes, meta data, for online recording etc) and improve messaging about NBN and how it can help*
By Nerds, For Nerds	*Improve functionality and usability of NBN family of projects and tools*
Social	*Build a coalition of the willing and prepare an online recording strategy for UK to build a bigger team*
Devolution	*Make the NBN a leader in managing and reporting on biodiversity data at any scale*

If you really want to get involved then come to our conference on Friday 21 November at the Royal Society, London or register for our NBN Gateway workshops the day before (Thursday 20 November at the Natural History Museum). Or attend one of our regional strategy workshops throughout the country.

Conclusions
In conclusion then, regarding the data deficits:

- Deficits exist but some are more important than others but most are treatable, teamwork is needed
- Deficits affect:
  - Conservation
  - Environment decision making
  - Research and recording
  - Education and learning
- Greater engagement is needed about biodiversity and importance of data and information
- The NBN will continue to build a critical mass of people that want to work together to deliver open, connected, usable data

Finally, what does that mean for the NBN? We will work to support members and strengthen the Network, We will work to increase use, visualisation and interoperability of data. We want to do this to inform state of environment reporting, conservation assessments (Red-listing os species and ecosystems), decisions about offsetting, and nature's services and natural capital programmes. We will work to increase engagement of the public, to inform communities about their neighbourhood wildlife. We will improve standards and data quality. And we will increase data diversity types and coverage.

That work will minimise deficits and will maximise innovation and positive environmental outcomes.

Thank you.