

The  
Alan Turing  
Institute

NATURAL  
ENGLAND

NBN

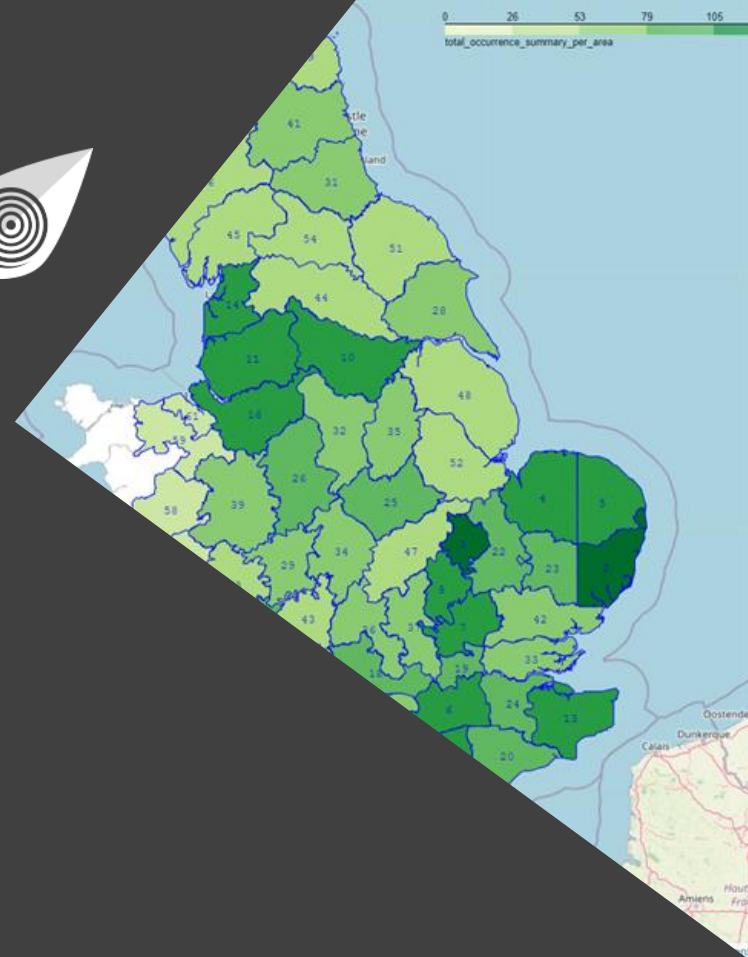


---

## Assessing NBN Atlas comprehensiveness and the impact of data providers: Insights from a Data Study Group

Simon Rolph

Data Scientist at  UK Centre for  
Ecology & Hydrology



---

**What is a data Study  
group?**

- Intensive '**collaborative hackathons**' which bring together organisations from industry, government, and the third sector, with talented multi-disciplinary researchers from academia
- Organisations act as Data Study Group '**Challenge Owners**', providing real-world problems to be tackled by small groups of researchers
- Researchers and a PI (Dr Silvia Liverani) develop **data science solutions**, presenting their work at the end of the week.

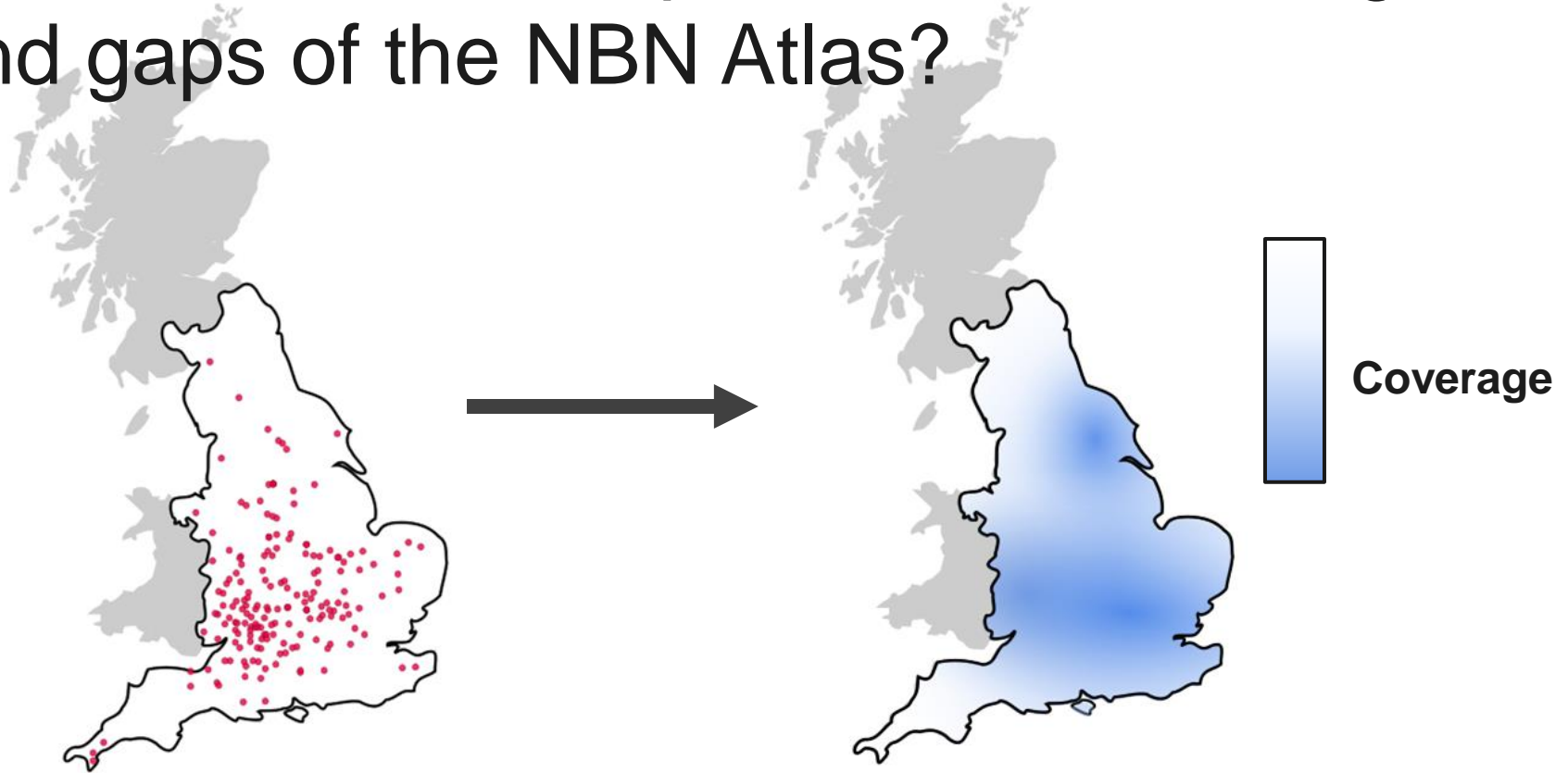


---

**What were our  
objectives?**

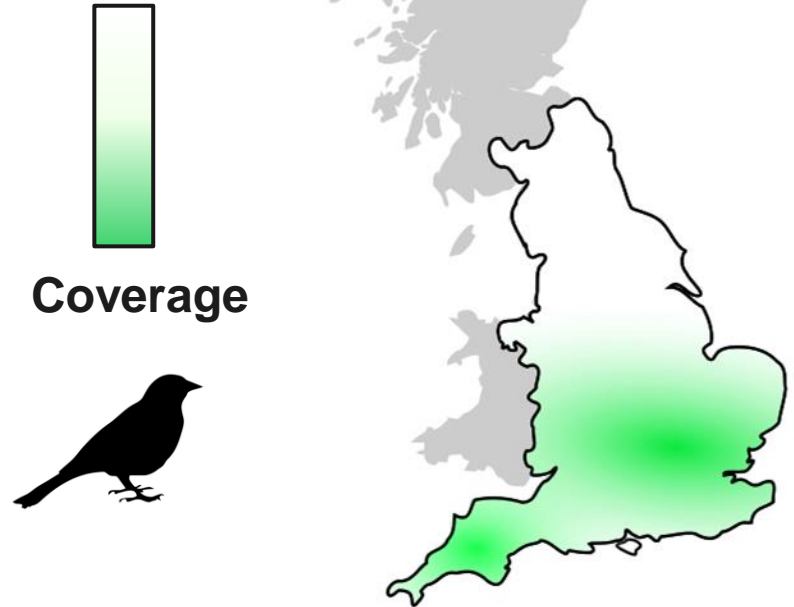
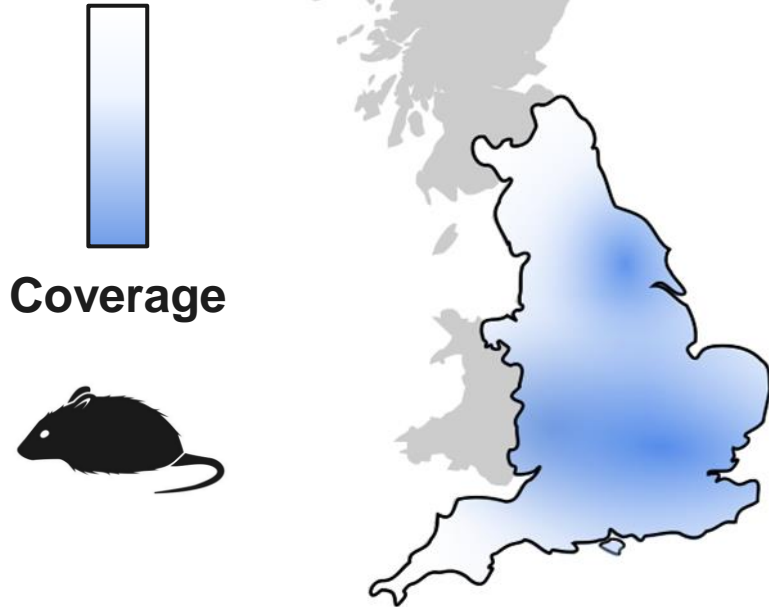
---

How can we best represent the coverage and gaps of the NBN Atlas?



---

How can we identify species that we do not have comprehensive coverage for?

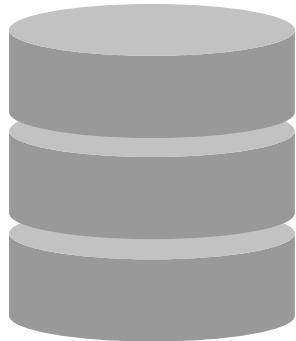


---

# How can we highlight the value that data providers bring to the Atlas?



**943** Priority species  
1970 - 2020 in England

A grey icon representing a database, consisting of three stacked cylinders. A white arrow points from the database icon towards the table.

#	Name	Common Name	Taxon	Year	Location
1	<i>Arvicola amphibius</i>	Water vole	Mammal	2019	52.9388° N 1.1981° W
2	<i>Apamea anceps</i>	Large Nutmeg	Moth	1975	51.7548° N 1.2544° W
3	<i>Passer domesticus</i>	House Sparrow	Bird	2001	50.3759° N 4.1396° W

⋮

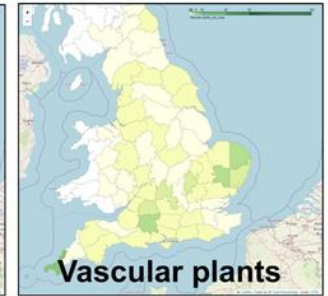
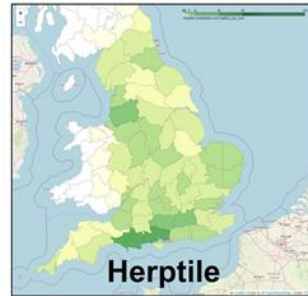
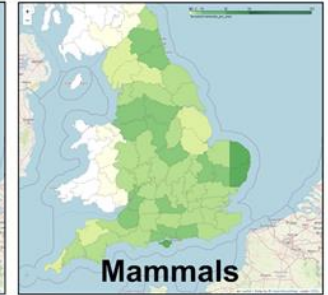
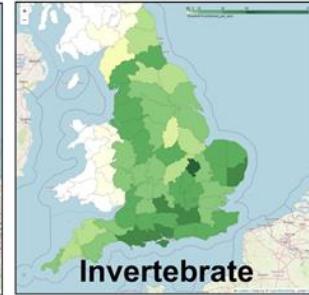
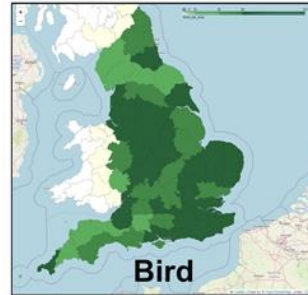
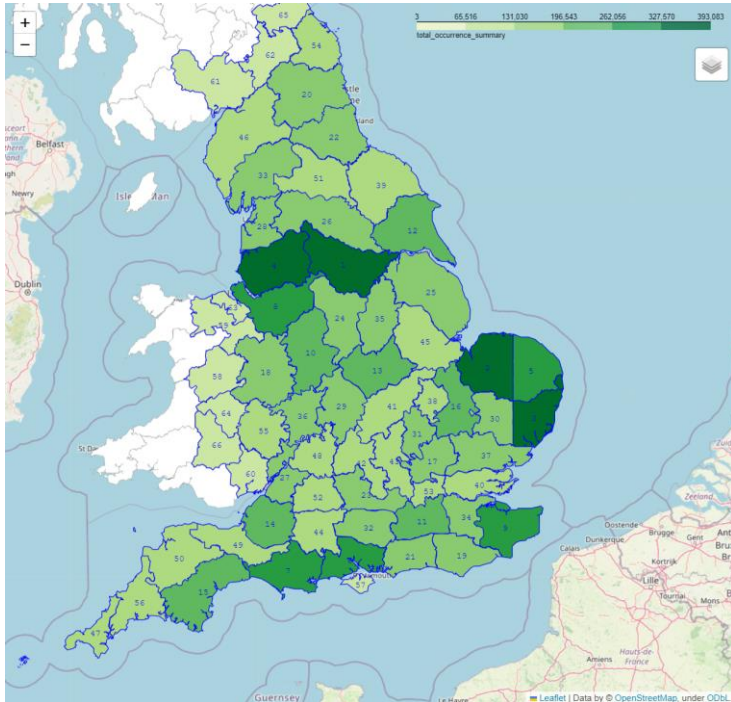
**10,319,464** Observations



---

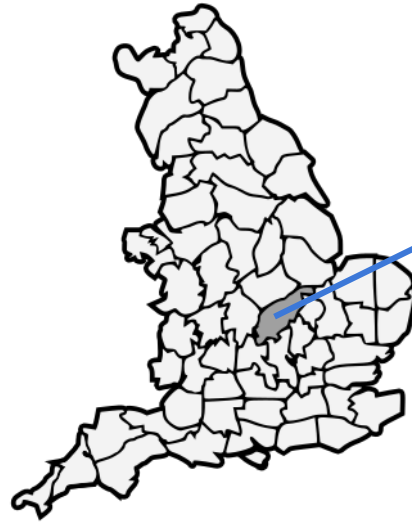
**How did we  
approach the  
challenge?**

# Getting to grips with the data: geographical and taxonomic biases for priority species



---

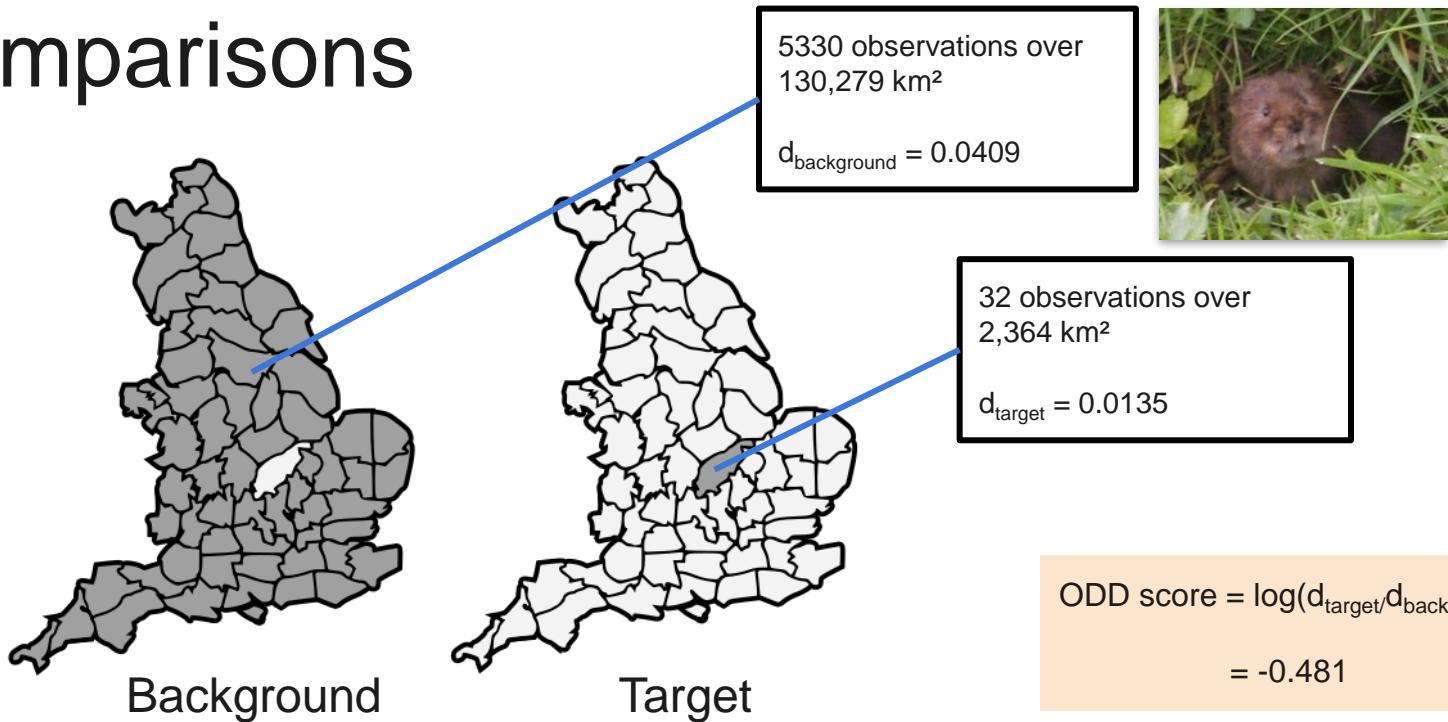
# Identifying gaps with reasonable comparisons



32 observations over  
2,364 km<sup>2</sup>  
 $d_{\text{target}} = 0.0135$

Target

# Identifying gaps with reasonable comparisons



# Occurrence density discrepancy

## Occurrence Density Discrepancy for Identifying Low Coverage Species (ODDFILCS)

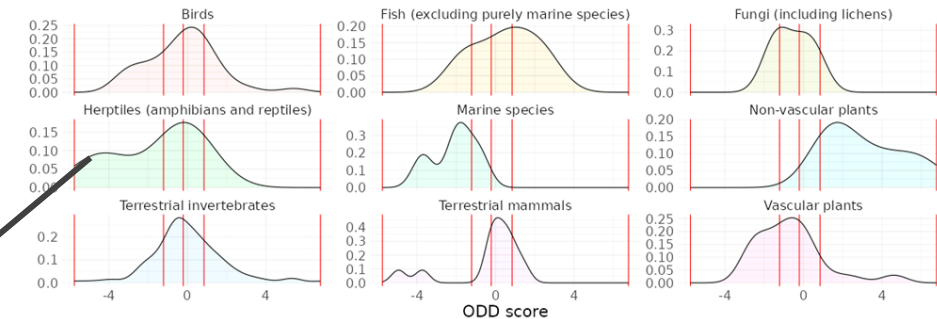
Vice County:

Cambridgeshire

Taxon Group:

All

User controls



Visualisation of ODD scores of species within taxon groups

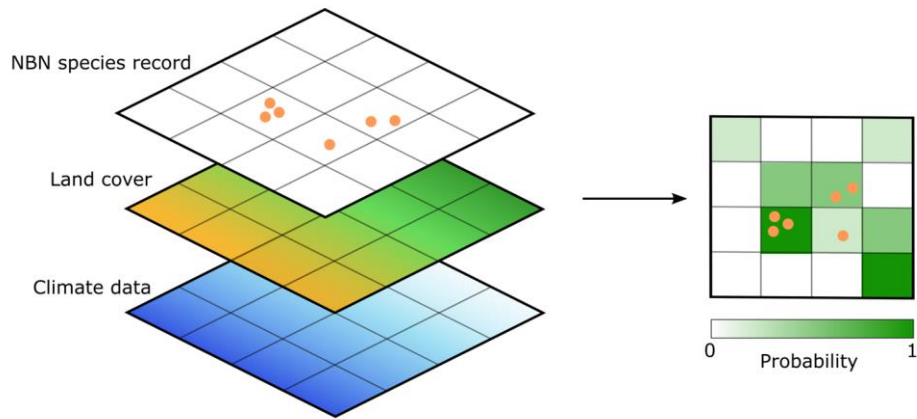
Show 10 entries

Search:

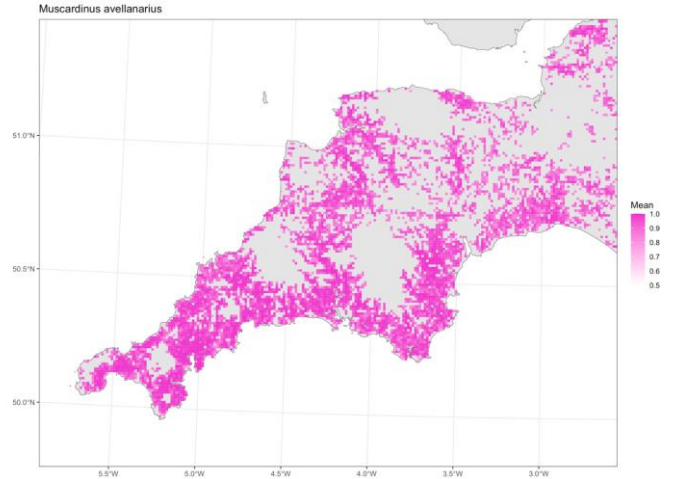
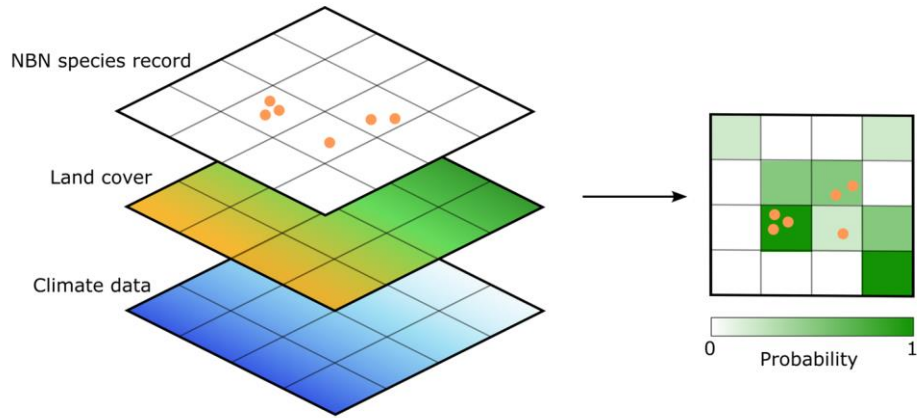
Species	Group	ODD score	Occurrences	Main provider	Quantile	% High res
1 Boloria selene	Terrestrial invertebrates	-5.8	1	UK Butterfly Monitoring Scheme	1	100
2 Sciurus vulgaris	Terrestrial mammals	-4.9	5	Norfolk Biodiversity Information Service	1	0
3 Limenitis camilla	Terrestrial invertebrates	-4.8	4	Cambridgeshire & Peterborough Environmental Records Centre	1	25
4 Vipera berus	Herptiles (amphibians and reptiles)	-4.8	3	Cambridgeshire & Peterborough Environmental Records Centre	1	0
5 Anguis fragilis	Herptiles (amphibians and reptiles)	-4	16	National Trust	1	75
6 Eugnorisma glareosa	Terrestrial invertebrates	-4	2	Cambridgeshire & Peterborough Environmental Records Centre	1	0
7 Lucanus cervus	Terrestrial invertebrates	-3.8	34	People's Trust for Endangered Species	1	100
8 Lepus timidus	Terrestrial mammals	-3.7	1	British Trust for Ornithology	1	100

Ranked list of species

# Environment-based gap identification



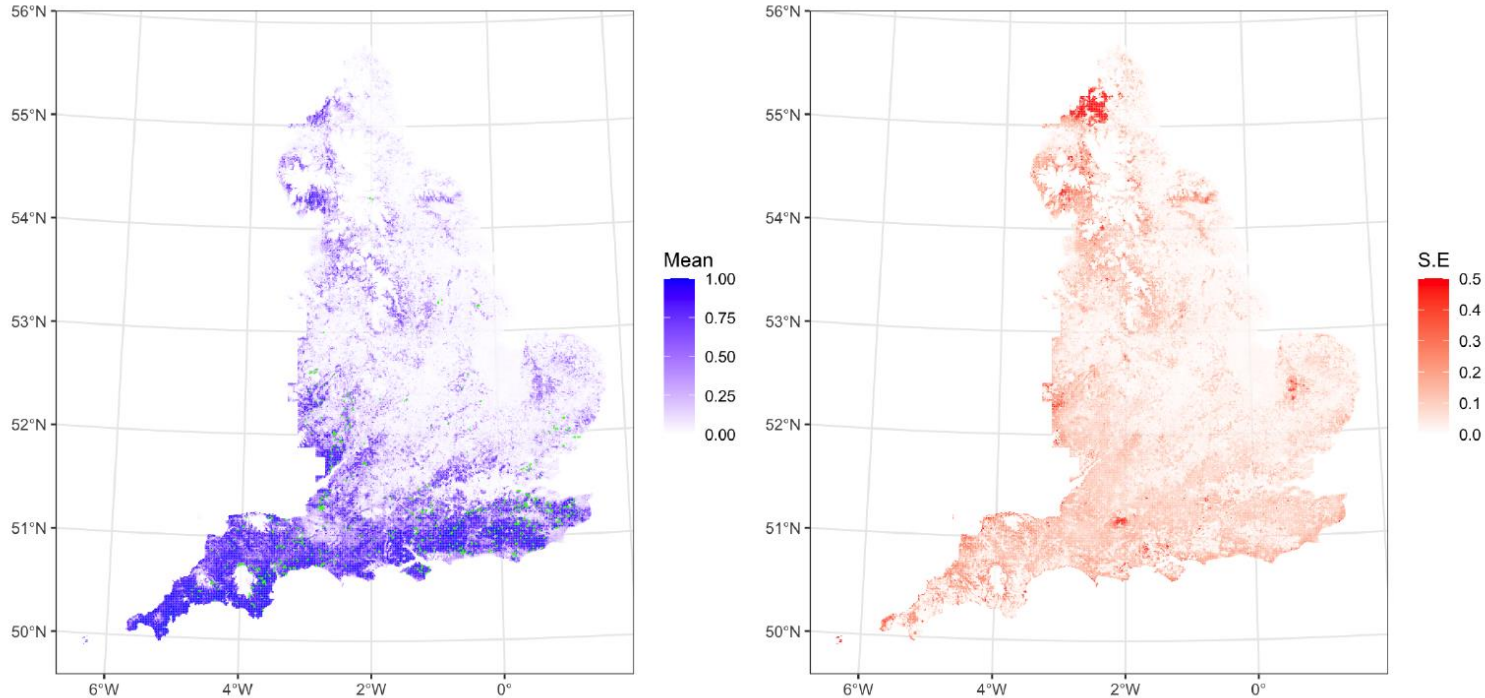
# Environment-based gap identification



- Species specific
- Able to highlight areas of interest



# Model-based adaptive sampling



Maps showing the 2018 probability of presence (purple), standard error (red) and actual records (green) for *Muscardinus avellanarius*

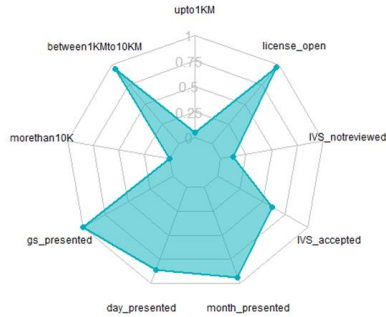


# We found 4 groups of data providers

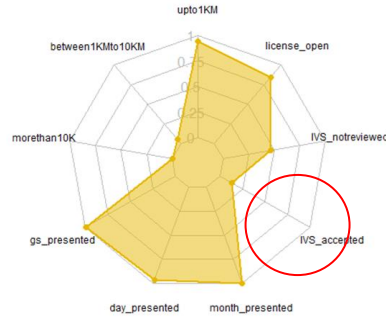
<b>Spot place precision:</b>	0-1 Km 1 Km to 10 Km More than 10 Km Not Provided
<b>Date information:</b>	Day_provided Month_Provided
<b>Data verification:</b>	Accepted Unconfirmed Not Reviewed
<b>Data usability:</b>	Open Shared



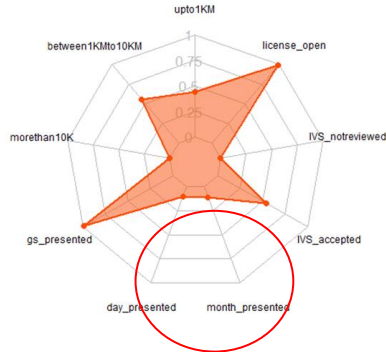
The regulars (30)



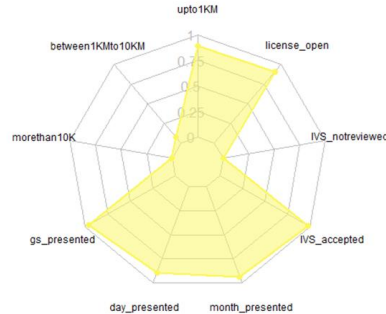
Unverified identification (17)



Incomplete date info (8)



Champions! (67)



# Helping data providers understand their data

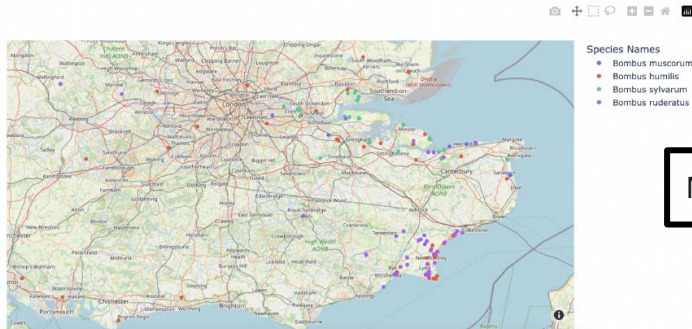
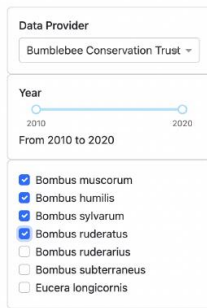
Prototype of a data provider dashboard

Bumblebee Conservation Trust

Total species: 7

Total records: 377

Average distance to HQ: 615.28 km



Map of records



Records over time

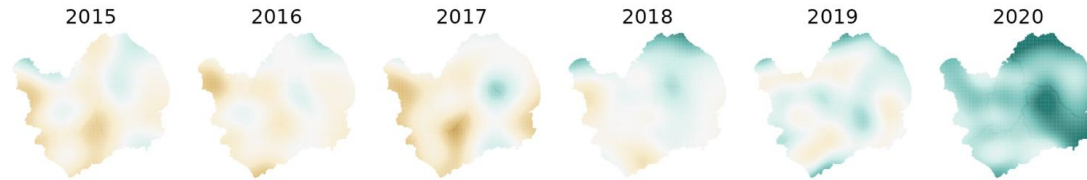
---

# Visualising the impact of data providers

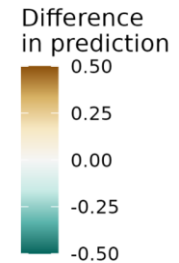
Can show influence of individual data providers by running species distribution models with and without their data

Example of hedgehog distribution in East Midlands:

PTES removed



BTO removed



---

# Main conclusions

- Gaps are multifaceted and complex
- Reasonable comparisons and prioritisation help identify gaps
- Combining with environmental is important for identifying gaps with models



## Reflections on a Data Study Group

- Time limited: successful ideas generation for future work, but many avenues unexplored
- Raised the profile of NBN Trust and the Atlas amongst data scientists
- Great learning experience for participants



---

# Report is online now

- Thanks for Natural England, the Alan Turing Institute and the NBN Trust for enabling this work



Search “*NBN Turing*”  
to find the report

The  
Alan Turing  
Institute

---

**Data Study Group  
Final Report:**  
National Biodiversity  
Network Trust

Spatiotemporal analysis of priority  
species records across England

22-26 May 2023

