Data driven futures: From stakeholder development to model development.

David Topping
David.topping@manchester.ac.uk

IAMA, UC Davis. 2023

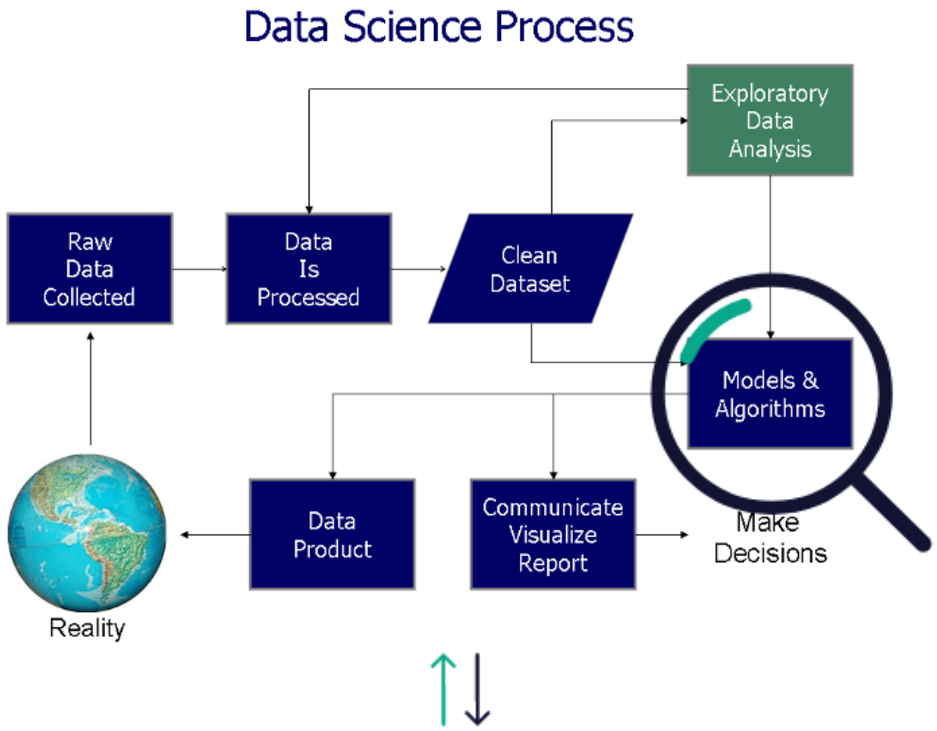

Data Driven Aerosol Science
NERC Digital Solutions Hub
Opportunities and Barriers


Data Driven Aerosol Science
NERC Digital Solutions Hub
Opportunities and Barriers

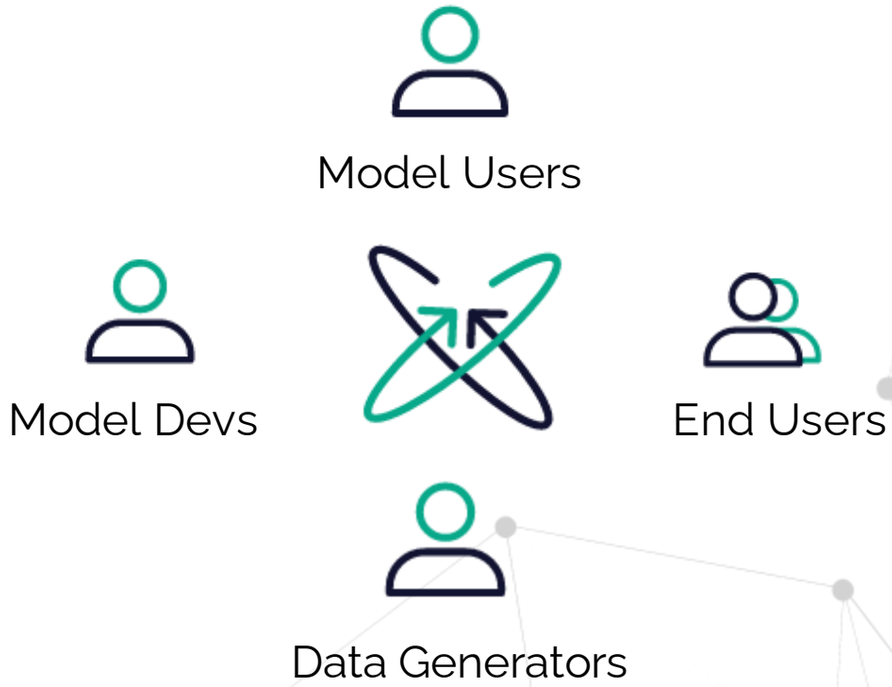
Data Driven Aerosol Science

Mechanistic frameworks that couple process to impact may not exist. *Machine learning* offer a potential route around that.

The *Data Science* ecosystem is critical to engage with a range of stake-holders, making effective use of data from a variety of sources - and creating a sustainable path to model development

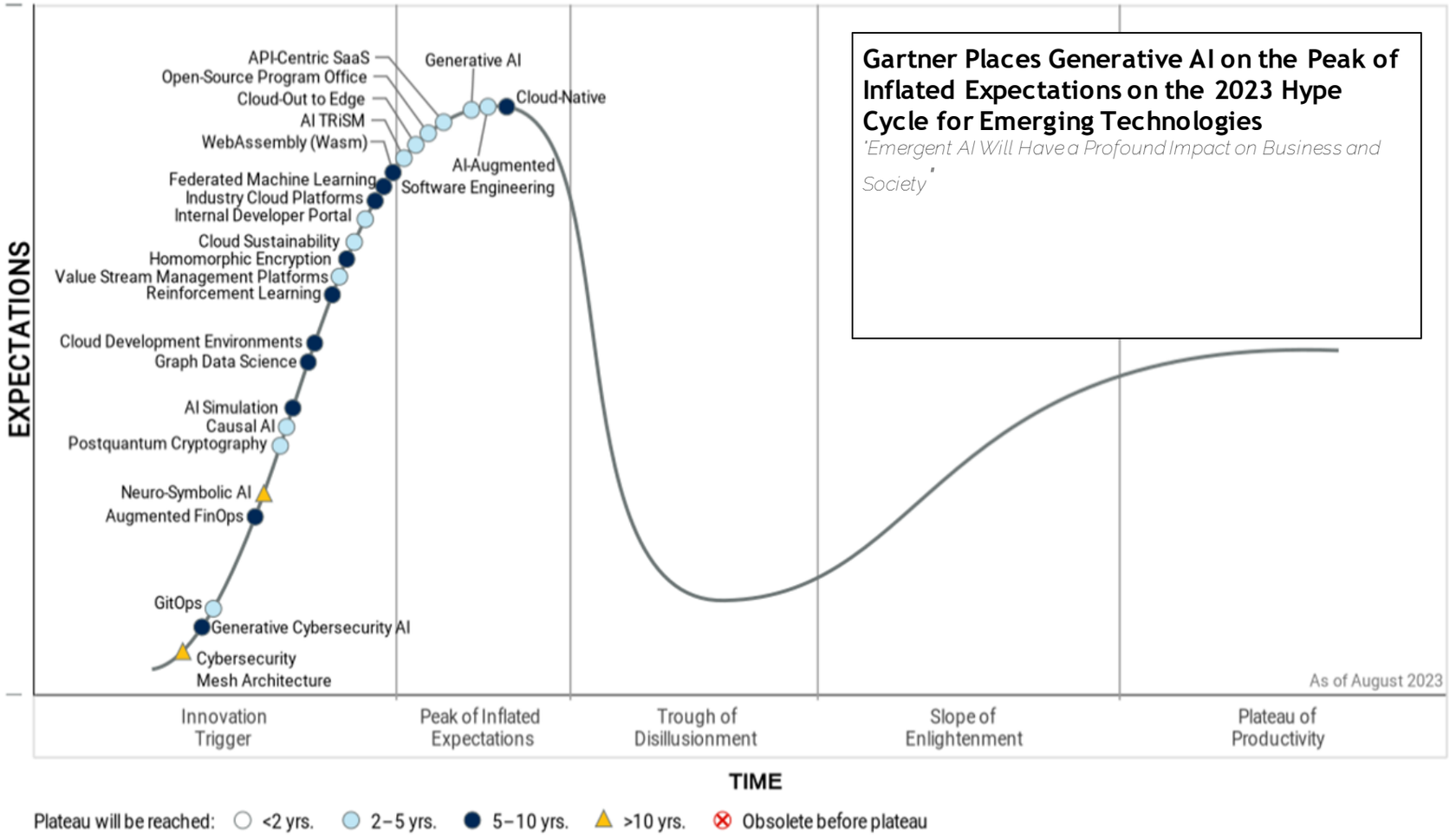


Ethics, Standards, Sustainability, Training...

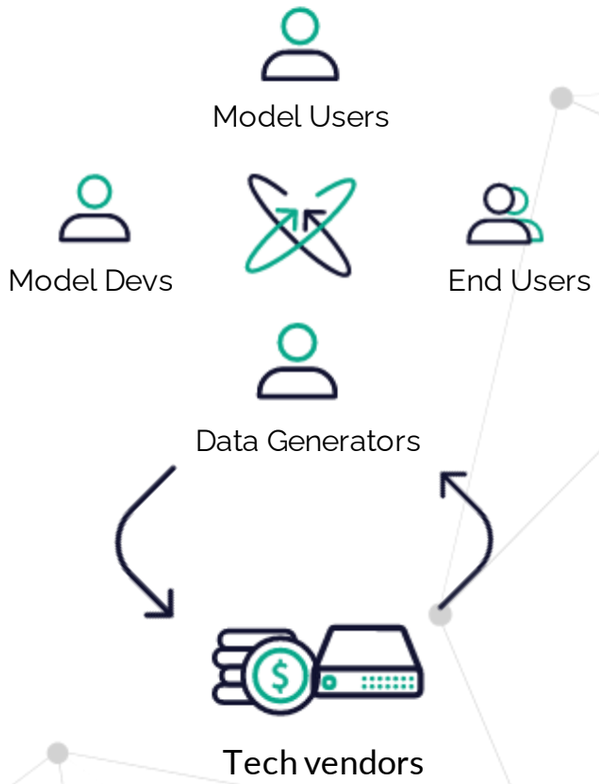


Doing Data Science, O'Reilly ISBN: 978-1-449-35865-5. Forbes-Gil Press-A Very Short History of Data Science-May 2013

Emerging technology moves into academia with increased speed

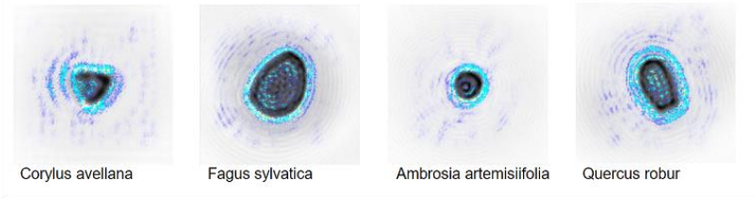


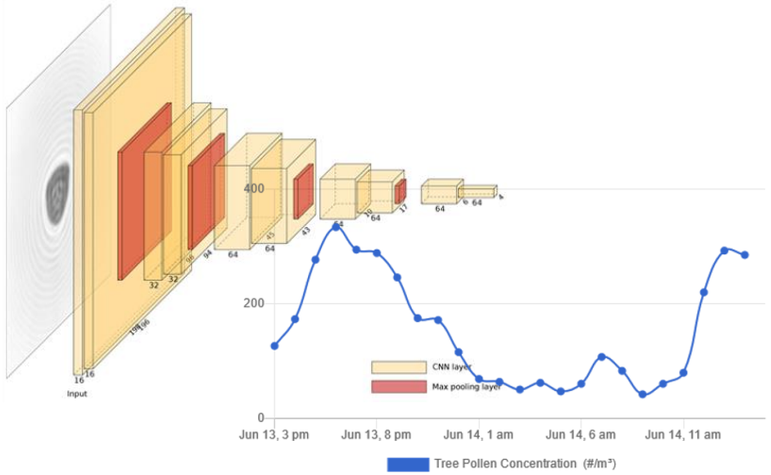
Gartner



From edge measurements....

e.g. realtime pollen

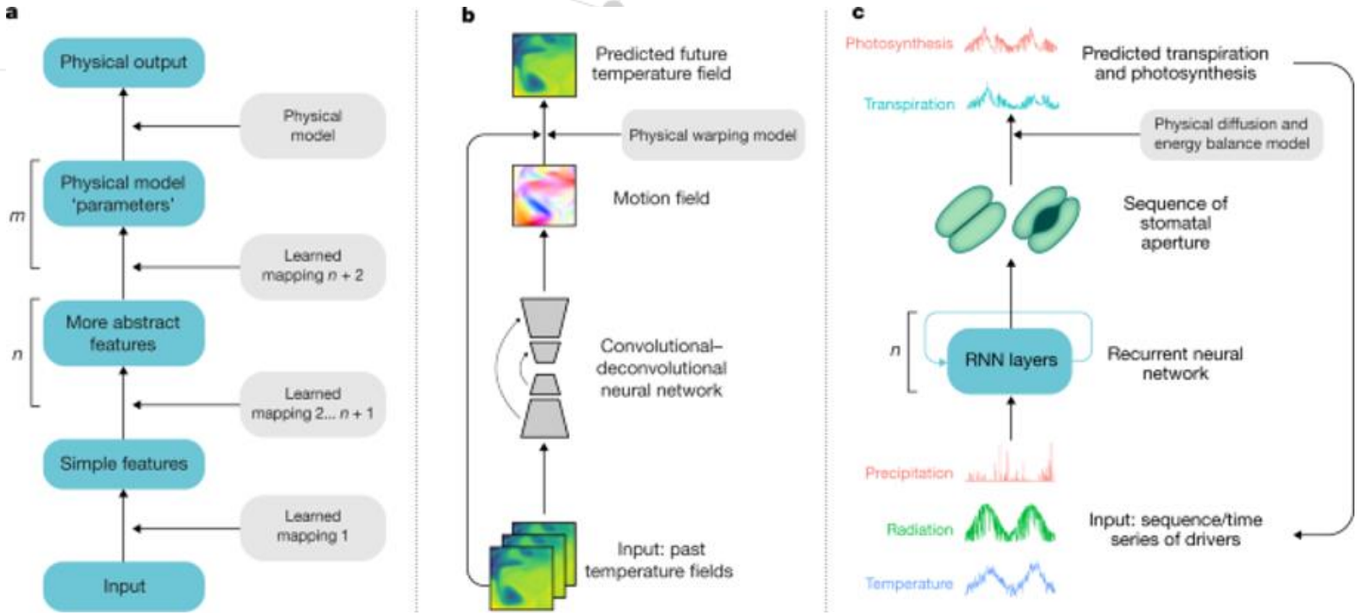




Sauvageat *et al Atmos. Meas. Tech., 13, 1539–1550, https://doi.org/10.5194/amt-13-1539-2020, 2020.*

to next generation models

e.g. improved nowcasting

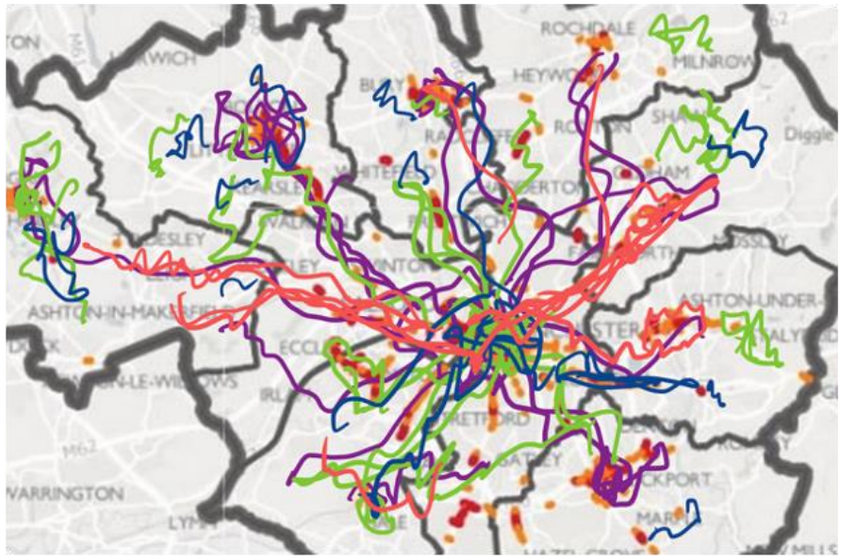


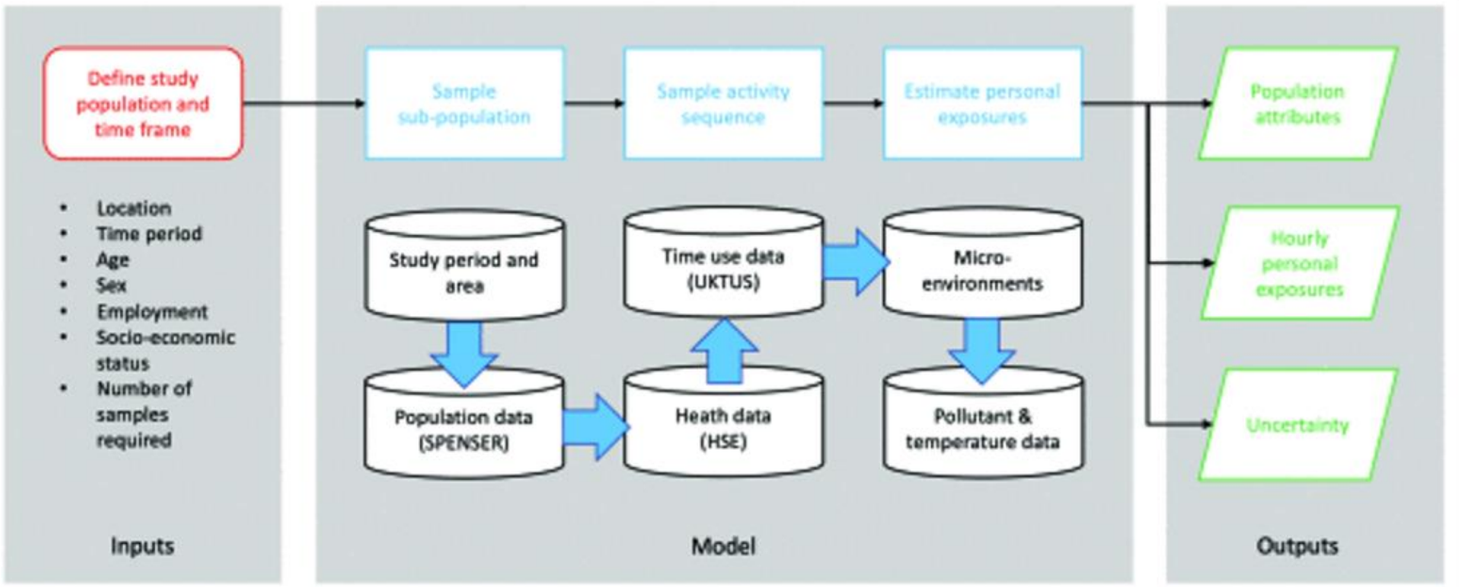
'combining machine learning with systems modelling is a key to the next level of earth system modelling' – Markus Reichstein

Reichstein, M., Camps-Valls, G., Stevens, B. et al. Deep learning and process

The path to Exascale includes hybrid process-ML models as part of the solution

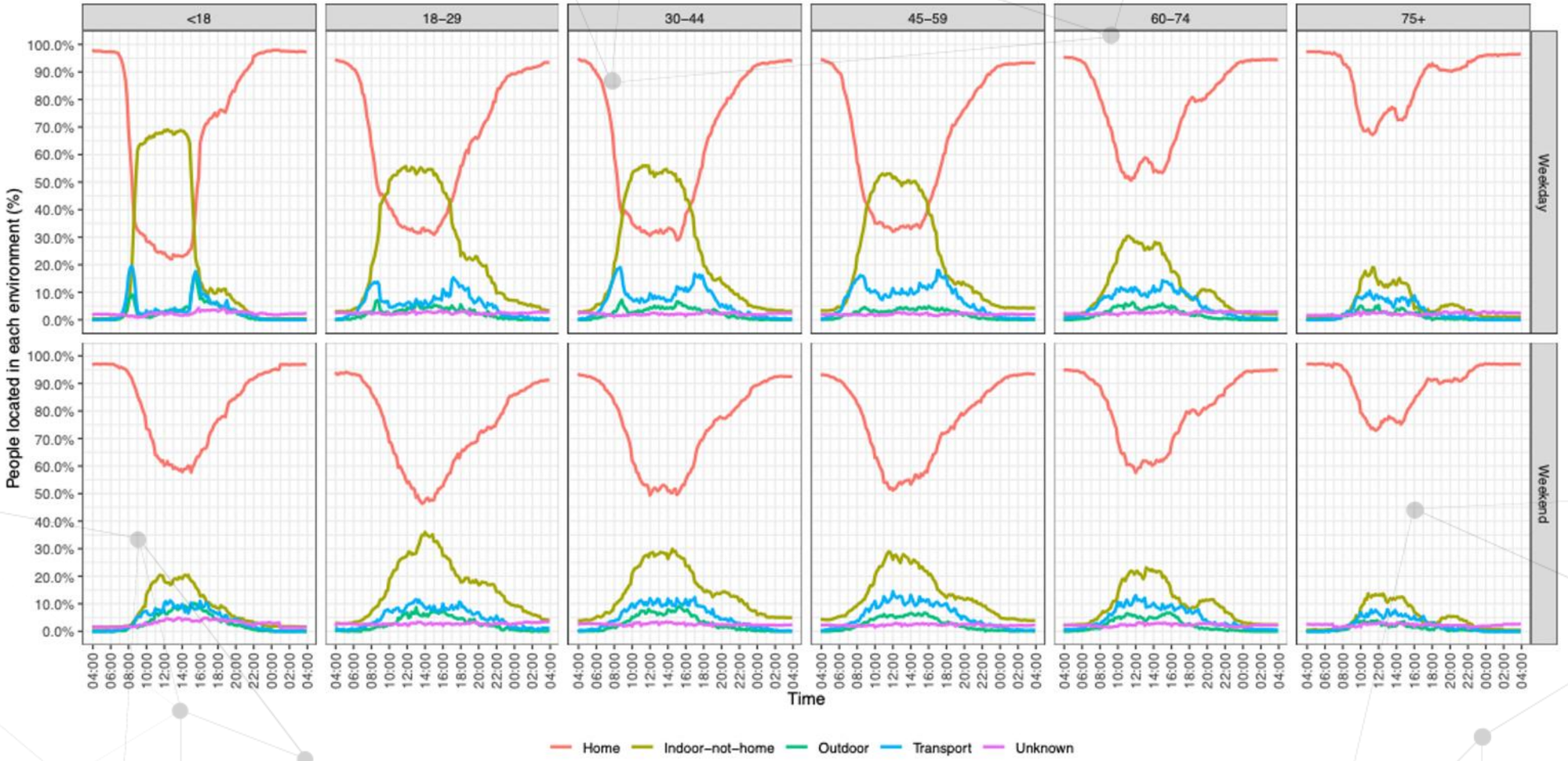
Crossing the health – environment interface using new personal exposure models





M. L. Thomas *et al.*, "A Data Integration Approach to Estimating Personal Exposures to Air Pollution," *2022 IEEE International Conference on Big Data (Big Data)*, Osaka, Japan, 2022, pp. 4551-4559, doi: 10.1109/BigData55660.2022.10020701.

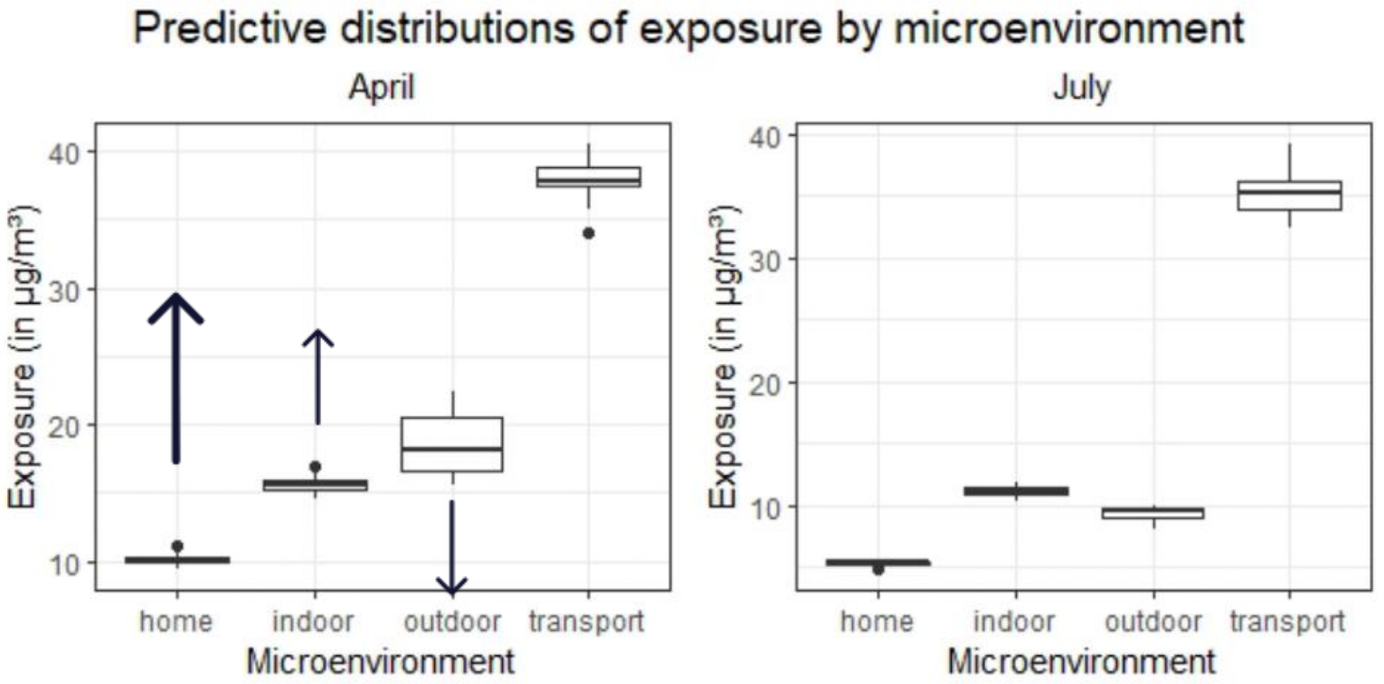
Crossing the health – environment interface using new personal exposure models



M. L. Thomas *et al.*, "A Data Integration Approach to Estimating Personal Exposures to Air Pollution," *2022 IEEE International Conference on Big Data (Big Data)*, Osaka, Japan, 2022, pp. 4551-4559, doi: 10.1109/BigData55660.2022.10020701.

Crossing the health – environment interface using new personal exposure models

- Raises a number of political conversations. Predictions based on **very limited data** in microenvironments.
- PM data indoors during cooking can become very high
- Decisions around urban mobility and greenspace already rolling forward.
- **How data is presented and its provenance is key…..**




Data Driven Aerosol Science
NERC Digital Solutions Hub
Opportunities and Barriers

DSH
Digital Solutions Hub

The **Digital Solutions Hub**

digital-solutions.uk

@nercdsh

NERC Digital Solutions Programme

•5 year £8m investment to build the Digital Solutions Hub
•40+ PBs data more discoverable **to non-academic users across** 5 data centres
•**British Oceanographic Data Centre** (marine)
•**Centre for Environmental Data Analysis** (atmospheric, earth observation, and solar and space physics)
•**Environmental Information Data Centre** (terrestrial and freshwater)
•**National Geoscience Data Centre** (geoscience)
•**UK Polar Data Centre** (polar and cryosphere).
•connecting
with **social**, **economic**, **health** and **environmental** data across the **whole of the UK**

Unique Property Reference Number (UPRN) for every addressable location across the UK. May **be any kind** of building, or it **may be an** object that **might not have a** 'normal' address – such as a bus **shelter.** **UPRN tagging** for geospatial data

https://github.com/sa-tre/satre-specification

Centre for Environmental Data Analysis
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

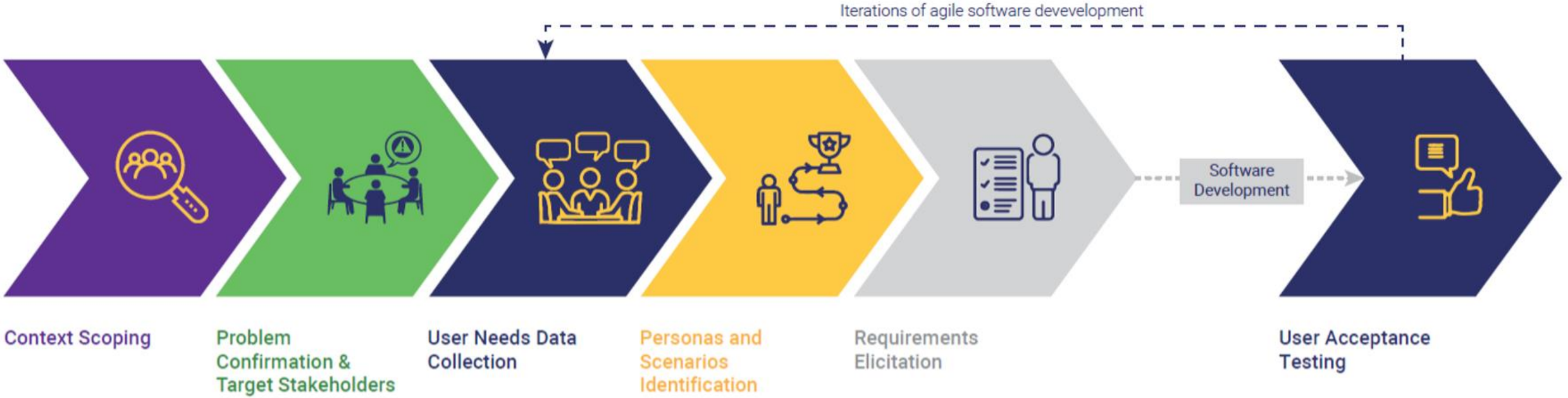
UK Centre for Ecology & Hydrology

British Antarctic Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

BGS British Geological Survey

BODC British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

User needs mapping approach



INTERACTIVE DATASET
273 data transactions
100 users
84 organisations

"If you build it, they might not come.."
Ask what users want first

User needs mapping approach

The
Analyst
Analysts monitoring the environment

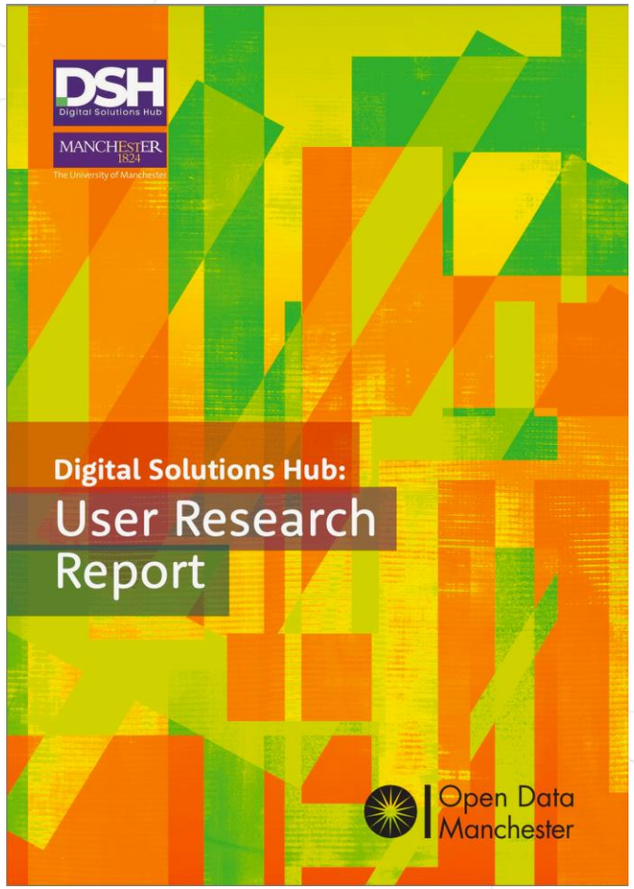
The
Investigator
Analysts answering questions with data

The
Author
Authors of monitoring frameworks

The Data
Specialist
GIS specialists and data support

The Data
Leader
Data leaders in organisations

The Data
Steward
Focusing on data quality and sharing



https://www.digital-solutions.uk/index.php/open-data-manchester-publish-the-report-from-our-user-engagement-workshops/

The landscape users operate in

•Data is held in **lots of places**.

•Data is not always held in formats, or on systems, that make it **easy to search for.**

•It's not always obvious what the purpose of different platforms is, and the variety of data they contain.

•It is hard to keep up with the '**sheer range**' of these.

•Some platforms are 'clunky' to learn and use with multiple clicks needed to get at data, or **arduous registration processes**.

•**Staff retention is hard** (e.g. cloud compute experts)

•Pressure to deliver with rapid government funding met with myriad of solutions

Key Requirements

•Users need sufficient **access to the data** to quality assure it, clean it and transform it into suitable format.

•**Reviewing a sample of a dataset** may help determine its suitability in an easier way.

•Users need to be able to **keep track of work and resources** they have done on datasets. [*we have introduced people in the same building!*]

•Users are keen to **avoid duplication of effort** – through sharing work they've done on datasets and accessing the work others have done.

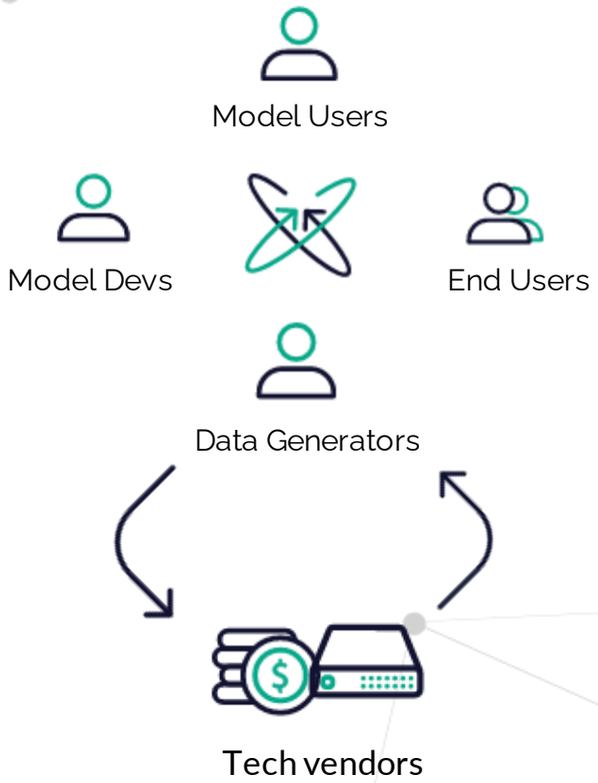
•A way of applying **suitable analysis** software to data.

•Allow users to **combine/link their own data with other data** they access from DSH as part of analysis.


Data Driven Aerosol Science
NERC Digital Solutions Hub
Opportunities and Barriers

Academic research developments progressing! Exciting times…
But...

- ‘Open data’ and 'open software' is no longer enough. Training is a heterogenous issue. What platforms will we be using in 5+ years? Will generative AI help? [see next slide]
- **Cultural challenges in data and software preservation and auditing around research data collection**
- There are software solutions that rely on governance and technical expertise. **We need to invest in people!**
- There are opportunities to build relationships with technology vendors. **What is the role of academia in this space?**
- Standards and regulation.



Turn negative perception into positive opportunity!

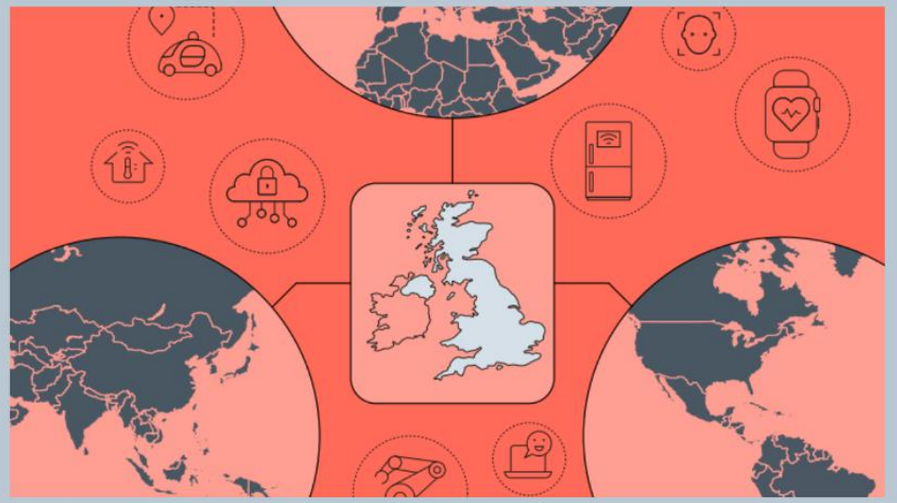
Report The future of regulation

Regulate to innovate

A route to regulation that reflects the ambition of the UK AI Strategy

29 November 2021

Reading time: 161 minutes



'**We must discover an ecology of technology**..' - James Bridle, 'Ways of Being'

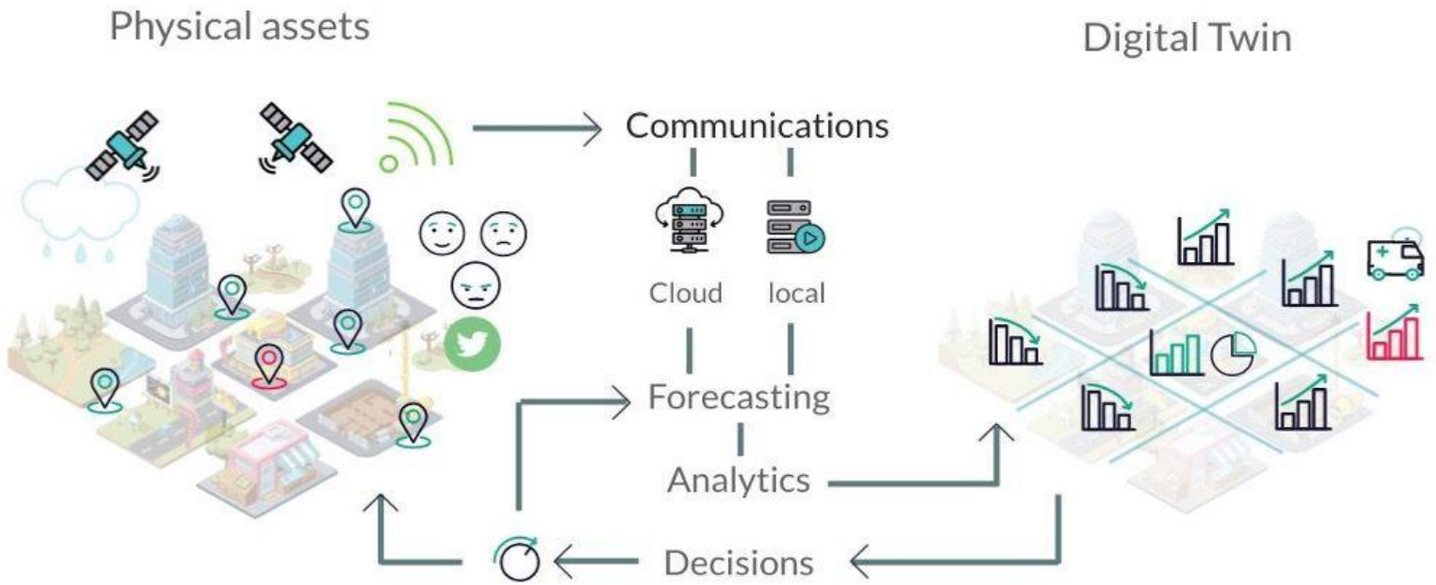
If we conducted an end user requirements exercise with ourselves (as model developers), what would we find?....**pinging IAMA**....

THANK YOU.

Please get in touch!

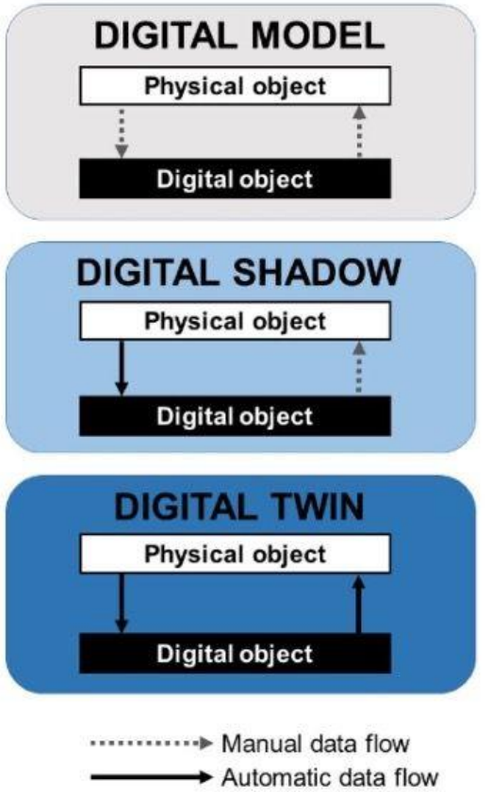
david.topping@manchester.ac.uk

A time and space for Digital Twins?

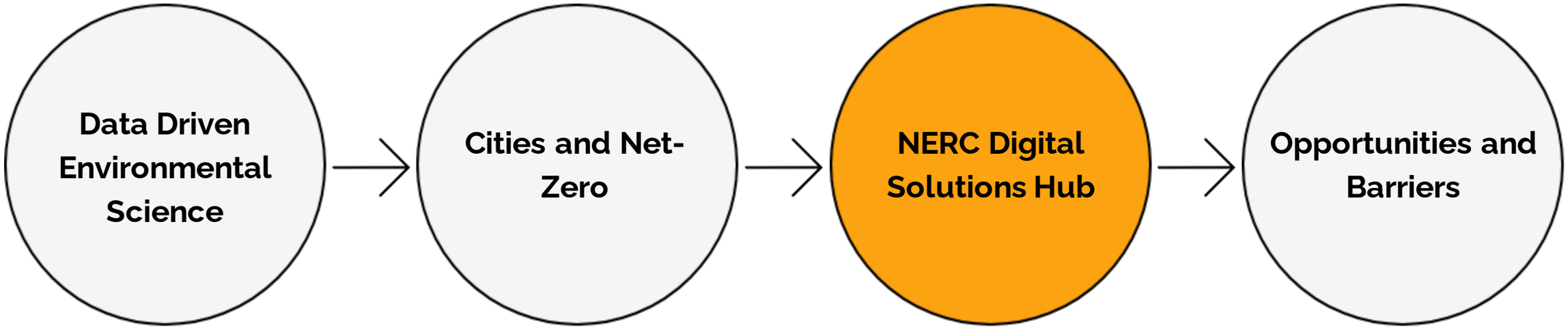


A Digital Twin could also allow a playback of multiple scenarios, simulating the potential impacts of various interventions before implementation in the real-world.

Topping D, Bannan TJ, Coe H, Evans J, Jay C, Murabito E and Robinson N (2021) Digital Twins of Urban Air Quality: Opportunities and Challenges. *Front. Sustain. Cities* 3:786563. doi: 10.3389/frsc.2021.786563

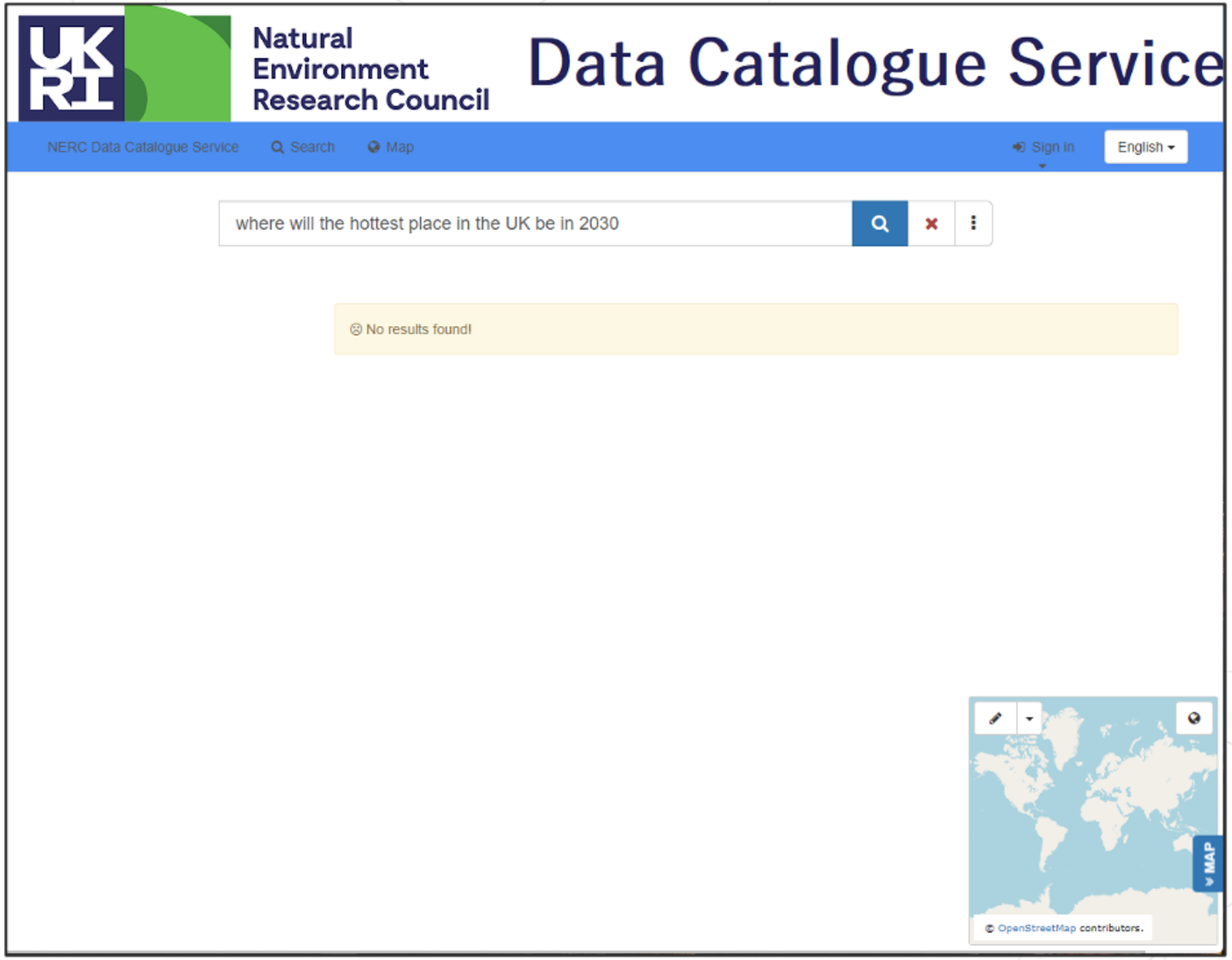


Digital models, shadows, and twins can be differentiated by the data flows between the physical and digital objects. (Image: Open Engineering)


Data Driven Environmental Science
Cities and Net-Zero
NERC Digital Solutions Hub
Opportunities and Barriers

"But we make data open already!"

Can you find UKCP18 data easily?

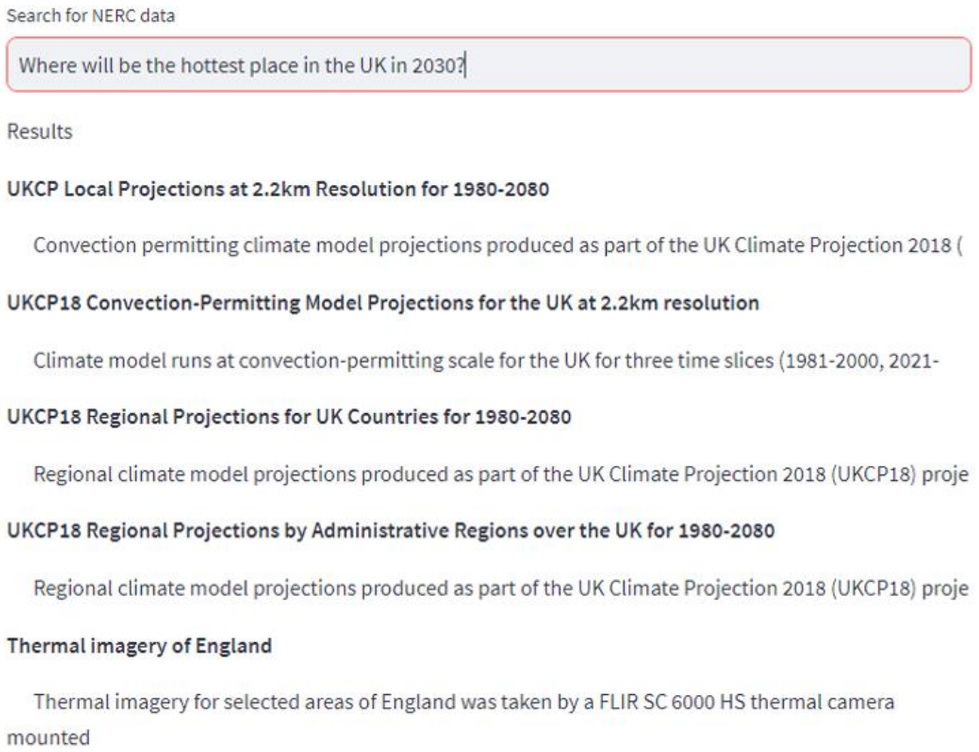


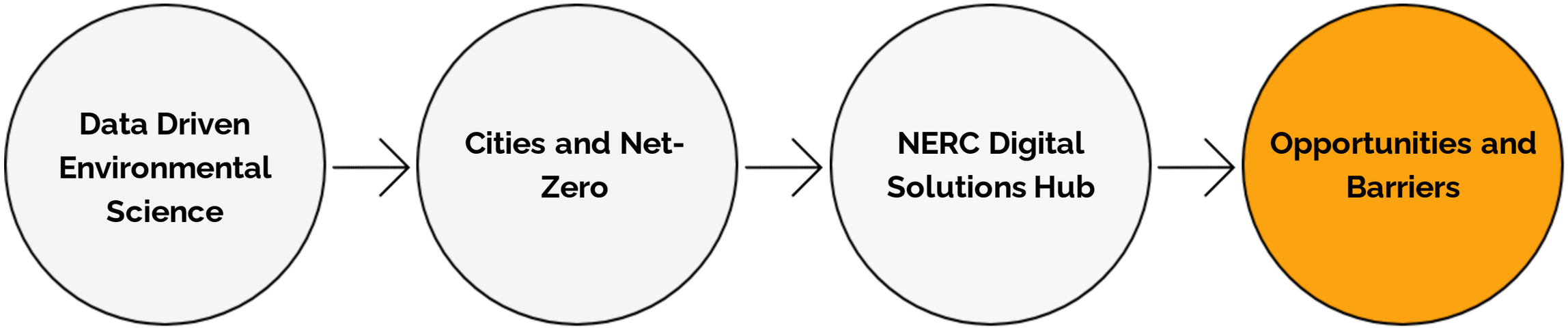
"But we make data open already!"

Can you find UKCP18 data easily?

Using a Large Language Model (LLM) we have trained it on all the NERC meta-data records.

Digital Solutions Hub




Data Driven Environmental Science
Cities and Net-Zero
NERC Digital Solutions Hub
Opportunities and Barriers