Office for National Statistics
Swyddfa Ystadegau Gwladol

Data Science Campus

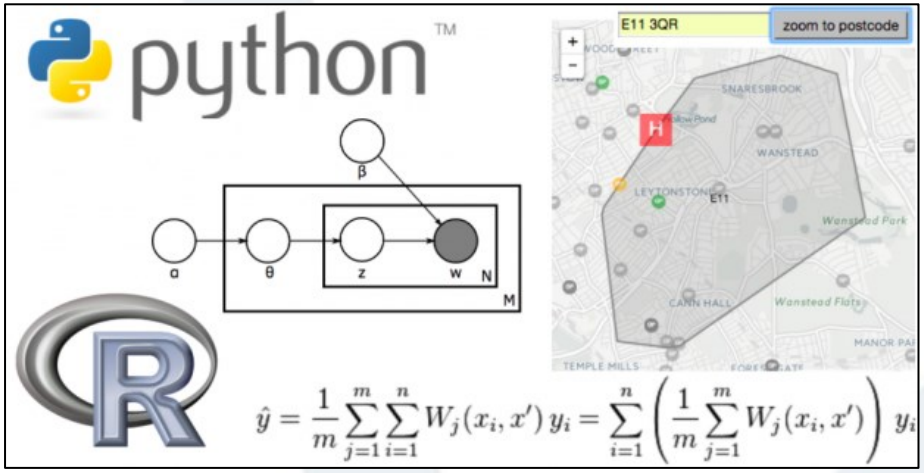
Campws Gwyddor Data

Welsh Statistical Liaison Committee, 16 February, 2017

Data Science & Big Data

Data science:

"..processes and systems to extract knowledge or insights from data.



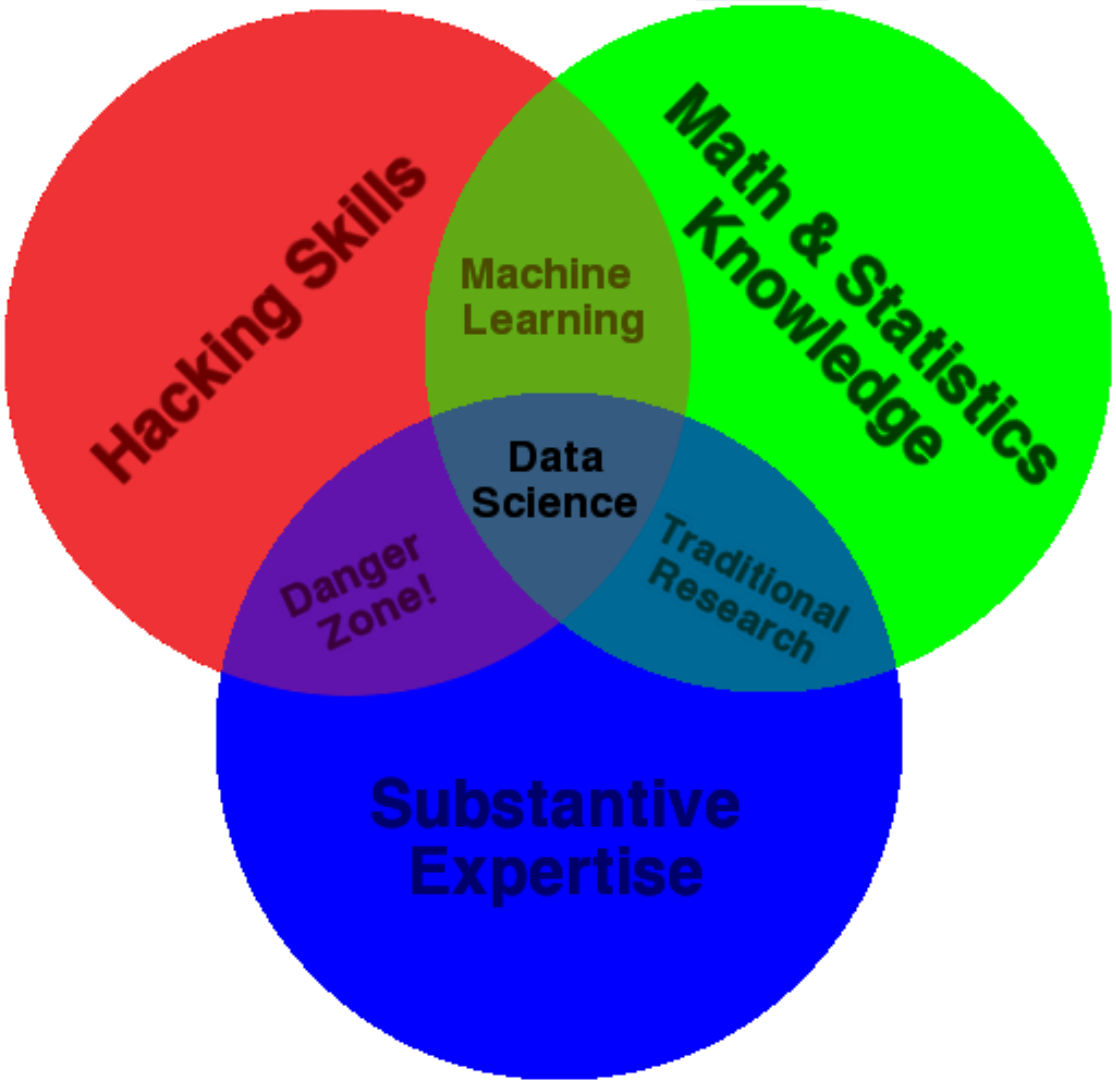
Big data:

"…so large and complex that traditional data processing applications are unable to deal with them.

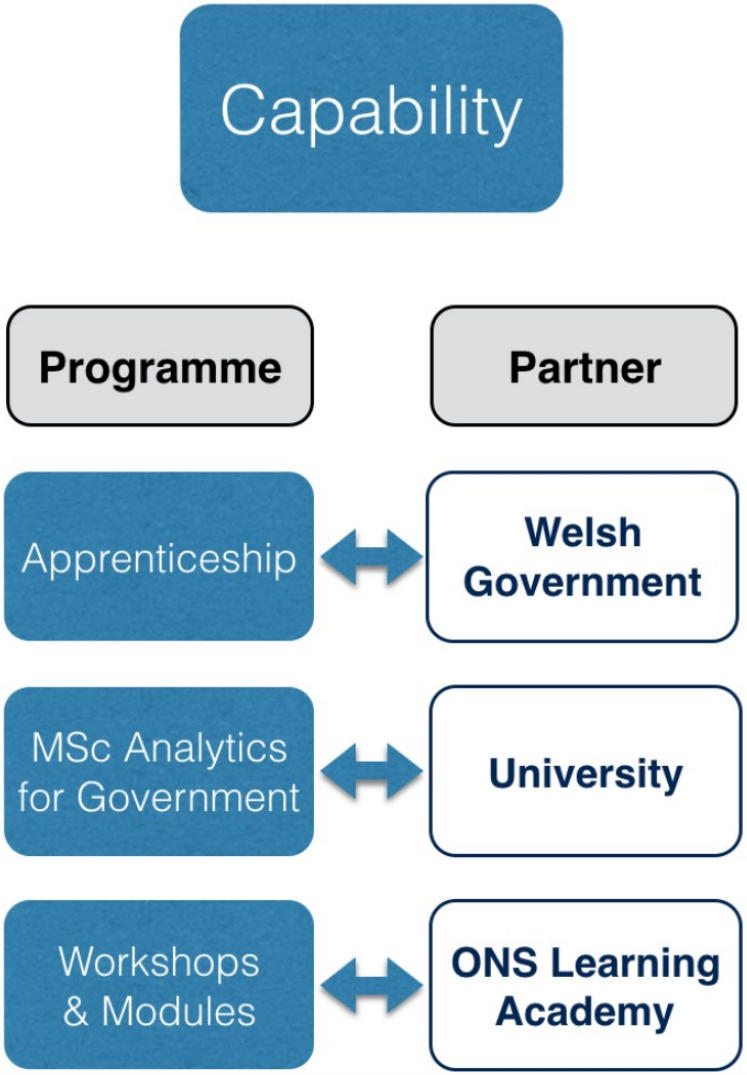
…inundates a business on a day to day basis.

..volume, velocity, variability.

Data Science:
a skill-set to understand a digital world



Bean Review: “Active Learning…



Apprenticeship in Data Analytics

- Two-year vocational programme
- Level 4 Diploma in Data Analytics
- Over 130 applications
- Eight apprentices from 5 December 2016

MSc Data Analytics for Government

- Dedicated Data Science pathway
- First intake in September 2017
- Multiple Academic partners
- Framework published 5th December 2016, partner applications closed 8th February
- Sponsoring 8 MSc’s in Data Science from October 2016

Data Science Accelerator

- 3-month programme
- Open to all Public Sector staff
- Run in conjunction with GDS and GO-Science
- Newport “hub” @ Data Science Campus

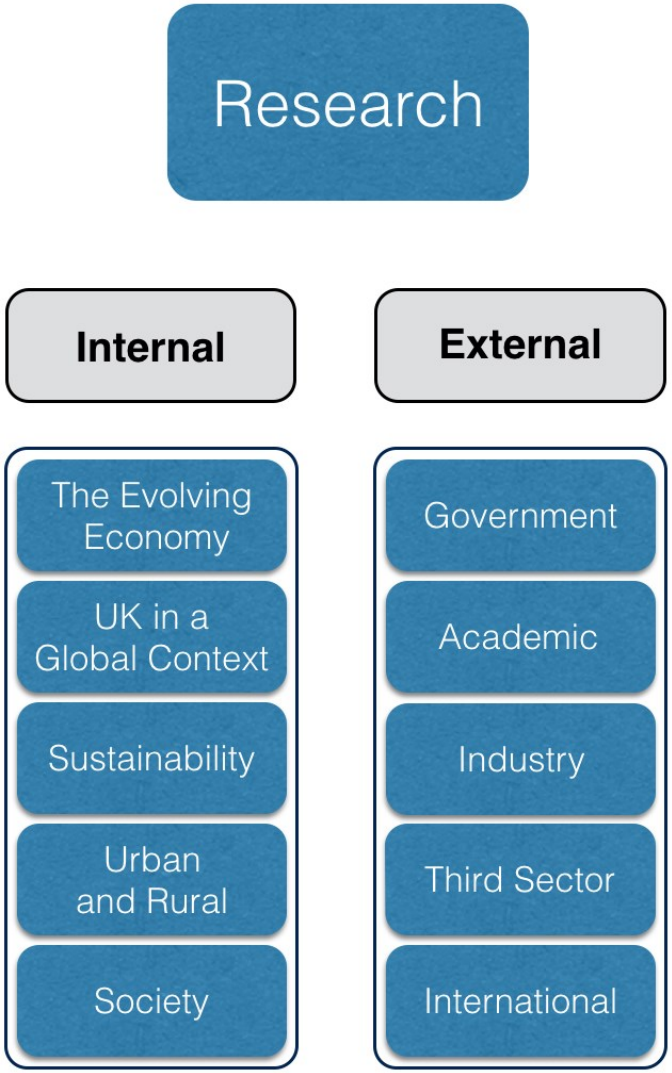
…and Experimentation"

Research Teams

- First research team launched in Sept 2016
- 25 FTE – new MD, Tom Smith arrived in January.
- 55 to 60 FTE Projected Headcount March 2018.

Partnerships

- MoUs agreed with multiple research partners including universities (e.g. Cardiff), research institutes (e.g. Alan Turing Institute) and international stats authorities (Stats Netherlands).
- Collaborations with national and devolved government including DEFRA, DCMS, DFID and Welsh Government.



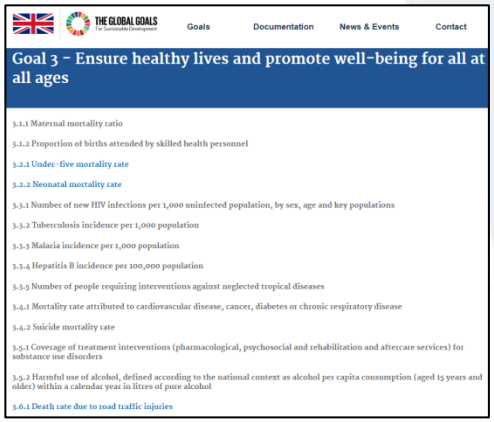
Data collection and dissemination tools

SUSTAINABLE DEVELOPMENT GOALS

Transforming Our World: the 2030 Agenda for Sustainable Development

Goal:

- Investigate how to build a data collection & dissemination system for SDG
- 6-week feasibility study: PoC completed using Github (built upon experience learned from the US Executive Office for the President)



Research outcome:

- Currently scoping 4-month projects for the SDG branch
- Future collaboration with US Executive Office for the President (EOP)
- Opportunity to reach out to developing countries: affordable open source solutions
- Cross-cutting value: reuse knowledge gained for other projects e.g. Cabinet Office online tool

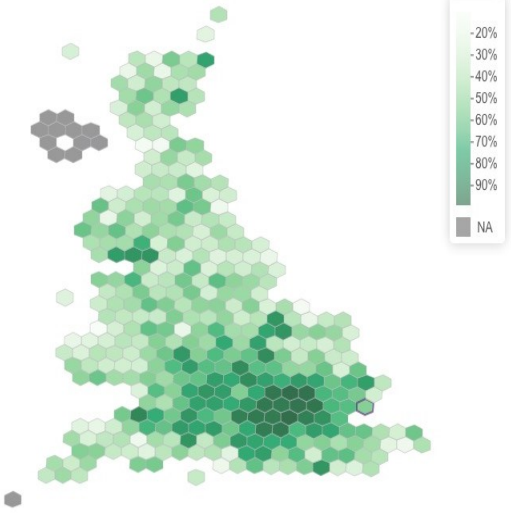
Industrial Risk/Resilience online tool

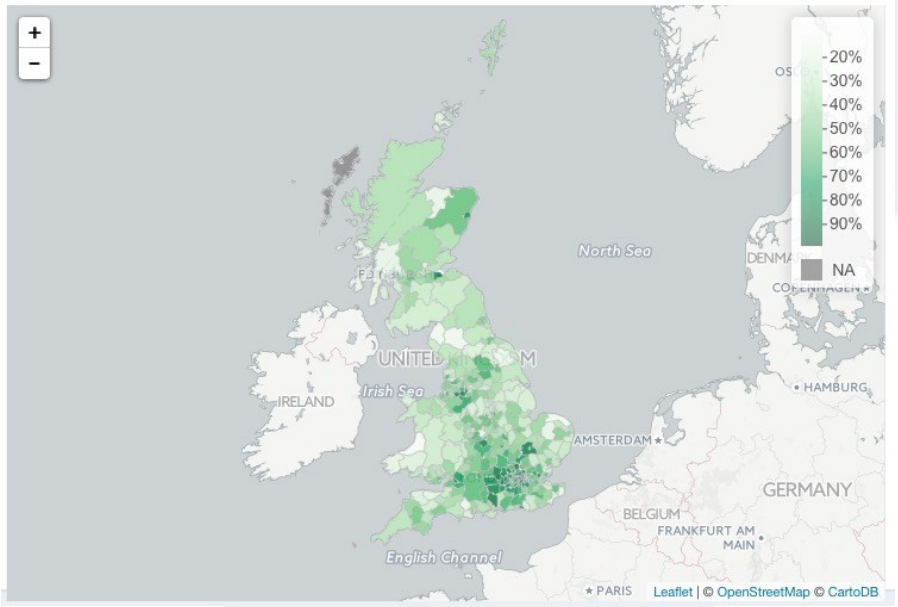
- Digital tool using R, Shiny and Leaflet to disseminate statistics on the resilience of employment at a local authority level.

- Hexagonal and geographical map of UK split into the local authorities, shaded by the value of the variable in question. Option to compare statistics between local authorities.

- Scatter plots of the two composite variables to show how the four quadrants in the graph indicate a different type of present and future situation for each local authority.

Cabinet Office





Estimate and explain tourism statistics

Our partners:

Department for Culture Media & Sport

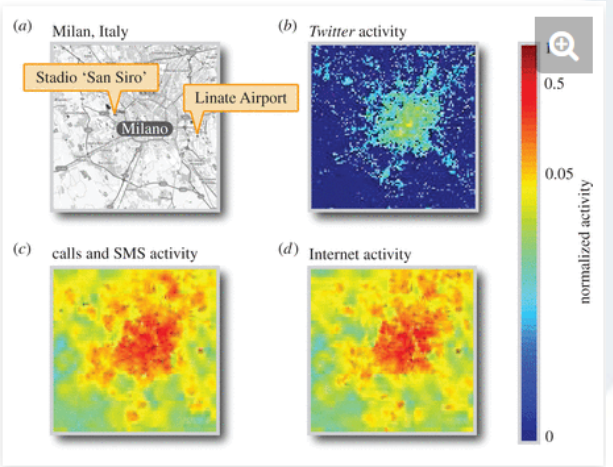
Office for National Statistics

wbs WARWICK BUSINESS SCHOOL

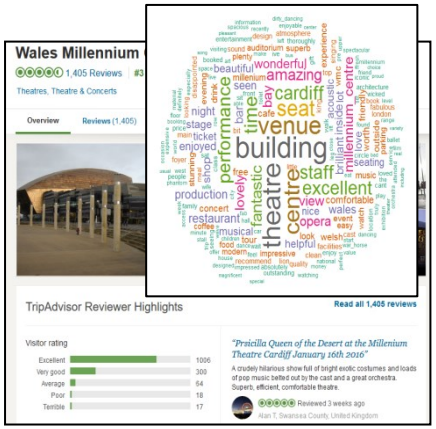
Centraal Bureau voor de Statistiek

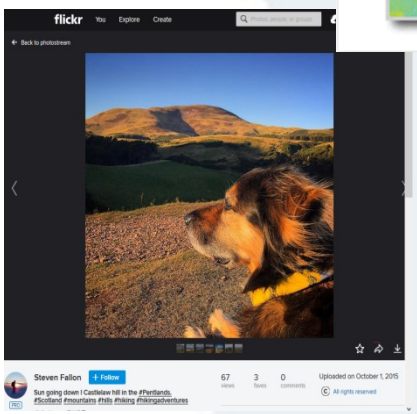
Research questions:

- Alternative data source for QA of the IPS
- Domestic travel trends
- Small area statistics, crowd size estimation
- Nationality based under-representation
- Google analytics web journey
- Draw insights: where do tourists go? Why?

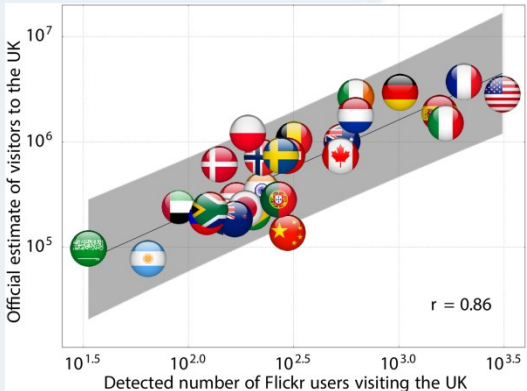


Visualisation of geo-located Flickr data





Machine Learning classification of photo tags



Economic and social impact of industries associated with fishing

defra Department for Environment Food and Rural Affairs

- The fishing industry supports significant levels of economic activity within many coastal communities.
- In collaboration with DEFRA, the Data Science Campus will better understand the importance of the fishing industry across the UK, particularly in deprived coastal communities.
- The initial phase will consolidate a range of data sets regarding measures such as employment, Gross Value Added (GVA) and fleet landings with IDBR data and spatially visualise on various scales.
- Better understand how the benefit of strong fishing activity propagates through the industry from production to services.

Measuring calorie consumption and obesity

- The Data Science Campus is collaborating in a project led by the ONS Health Division. Partners across the GSS, industry and academia:

defra Department for Environment Food and Rural Affairs

THE BEHAVIOURAL INSIGHTS TEAM IN PARTNERSHIP WITH Cabinet Office

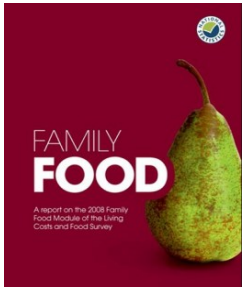
Public Health England

Department of Health

UNIVERSITY OF OXFORD

- **Background of the research:**
  - Data on body weight and on calorie consumption diverge more and more over time
  - Previous work by the Behavioural Insight Team: under-reporting is the most plausible explanation. The decline in physical activities does not cause significant weight gain
- **However:** our research will not rule out any hypotheses. Apply data science techniques to draw insights from various data sources and linking information

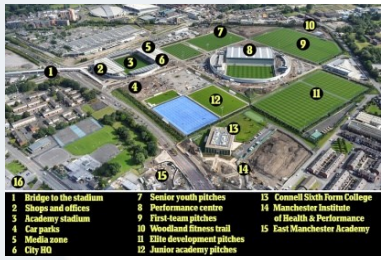
Supermarket point of sale data



Living cost and Food survey

Data on food wastage

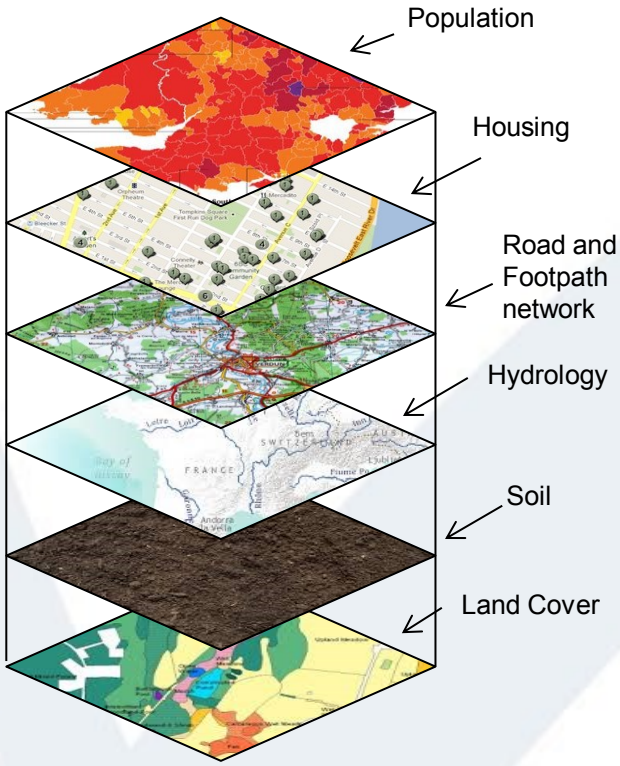
Oxford University Centre for Time Use Research: activity diary + fitbit + camera



Map of infrastructure and sport facilities

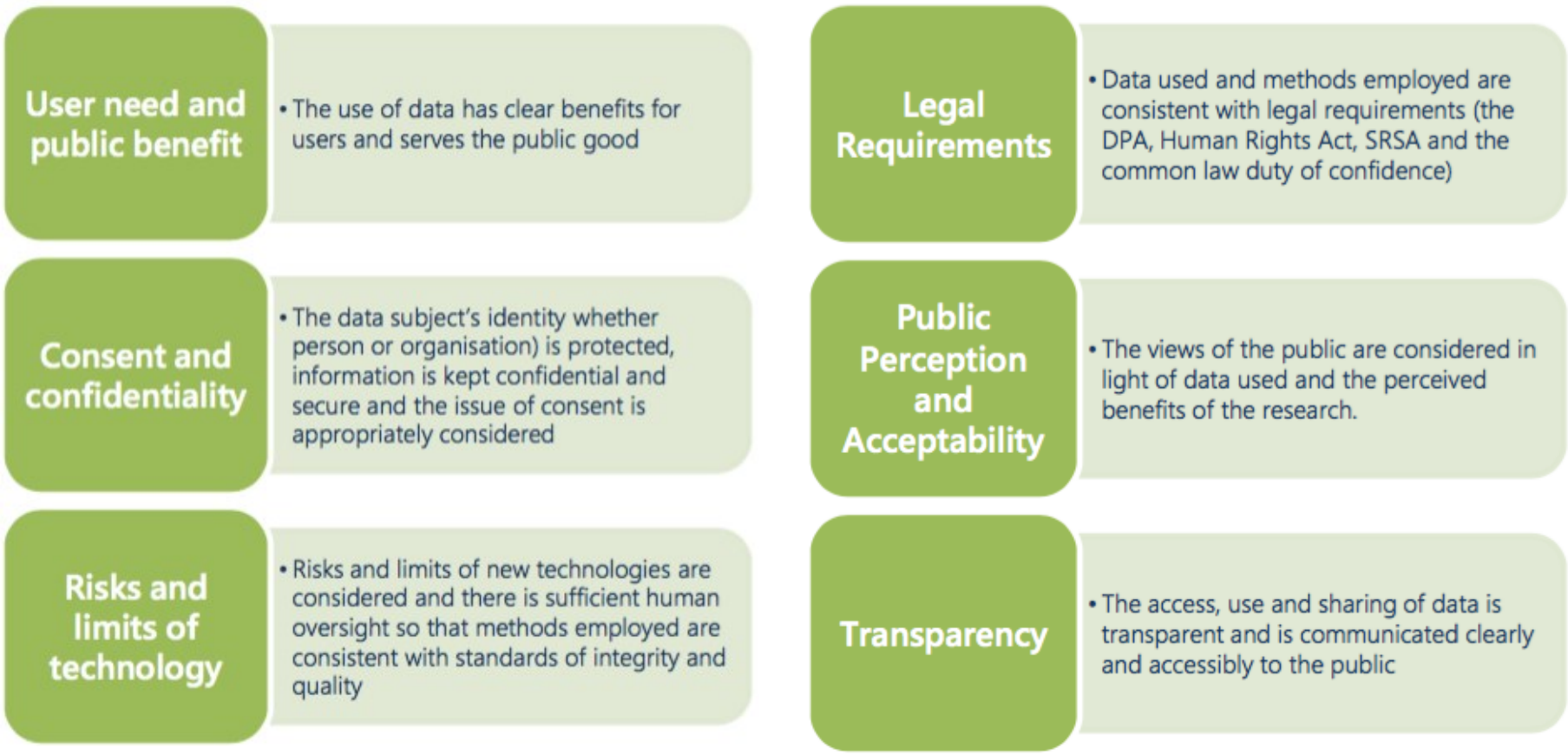
Quantifying the urban forests

- **Overview:** this project fits into the wider context of the Natural Capital Accounting effort led by the ONS WISE Division, DEFRA, the Environment Agency, Natural England. Potential collaboration with Welsh Government



- **What is the urban forest?**
  - All trees in the urban realm: public and private spaces
  - Along linear routes and waterways
  - In amenity areas
  - green infrastructure and the wider urban ecosystem
- **Benefits of trees/canopy:** energy saving, improving air quality, storm water attenuation, shade and cooling, storing carbon, biodiversity and habitat, aesthetic/landscape screening, mental health
- **Project scope:** utilise machine learning and machine vision techniques to create a UK wide urban tree inventory covering (but not limited to):
  - Tree counts
  - Height, canopy width, age, biomass..
  - Species
  - Context (public vs private land etc)
  - Individual tree indicators (accessibility, energy saving etc.)

Data Ethics and the Public Good



National Statistician's Data Ethics Advisory Committee (NSDEC)

Data Science Campus:

www.ons.gov.uk/datasciencecampus

datasciencecampus@ons.gov.uk

peter.j.fullerton@ons.gov.uk