Building Data Science Capacity at ONS and Beyond

Monday 29th April, 2019

Tom Smith
Managing Director
UK Data Science Campus

Tom Wilkinson
Head of MI and Analytics
DFID

Ceri Regan
Academic Manager
UK Data Science Campus

Overview

We will present, as follows:

1. Tom Smith - MD, Data Science Campus, ONS, UK
   - The UK Data Science Campus journey, a bit of history

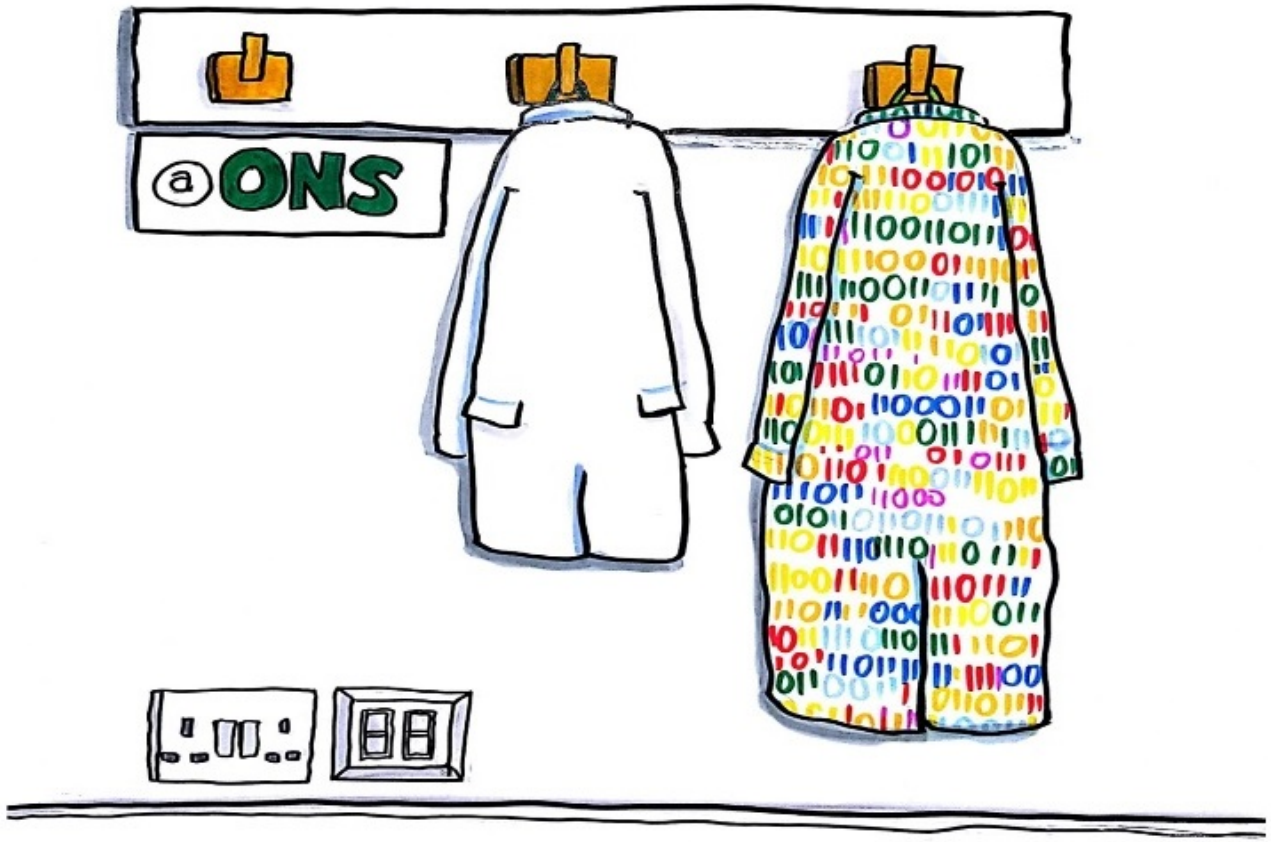
2. Tom Wilkinson – Head of MI & Analytics, DFID, UK
   - Data Science in UK Gov, collaboration, international outreach

3. Ceri – Academic Manager, Data Science Campus, ONS, UK
   - Our experience of building data science capability capability, the work we are doing with Rwanda/UNECA

If time – Discussion or Q&A session

UK Data Science Campus – Mission and Story

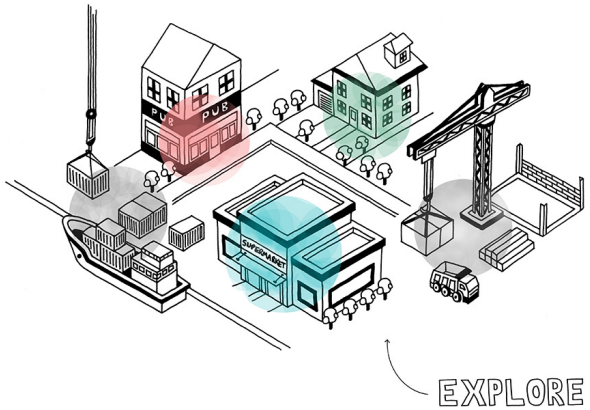
Tom Smith, @_datasmith
Director, ONS Data Science Campus



Data Science Campus

web: datasciencecampus.ons.gov.uk
email: datasciencecampus@ons.gov.uk
twitter: @DataSciCampus

Office for National Statistics



Economy

GDP
Inflation
Labour market
+++

People

Population
Census
Incomes
+++

World

Trade
Sustainable Development Goals
+++

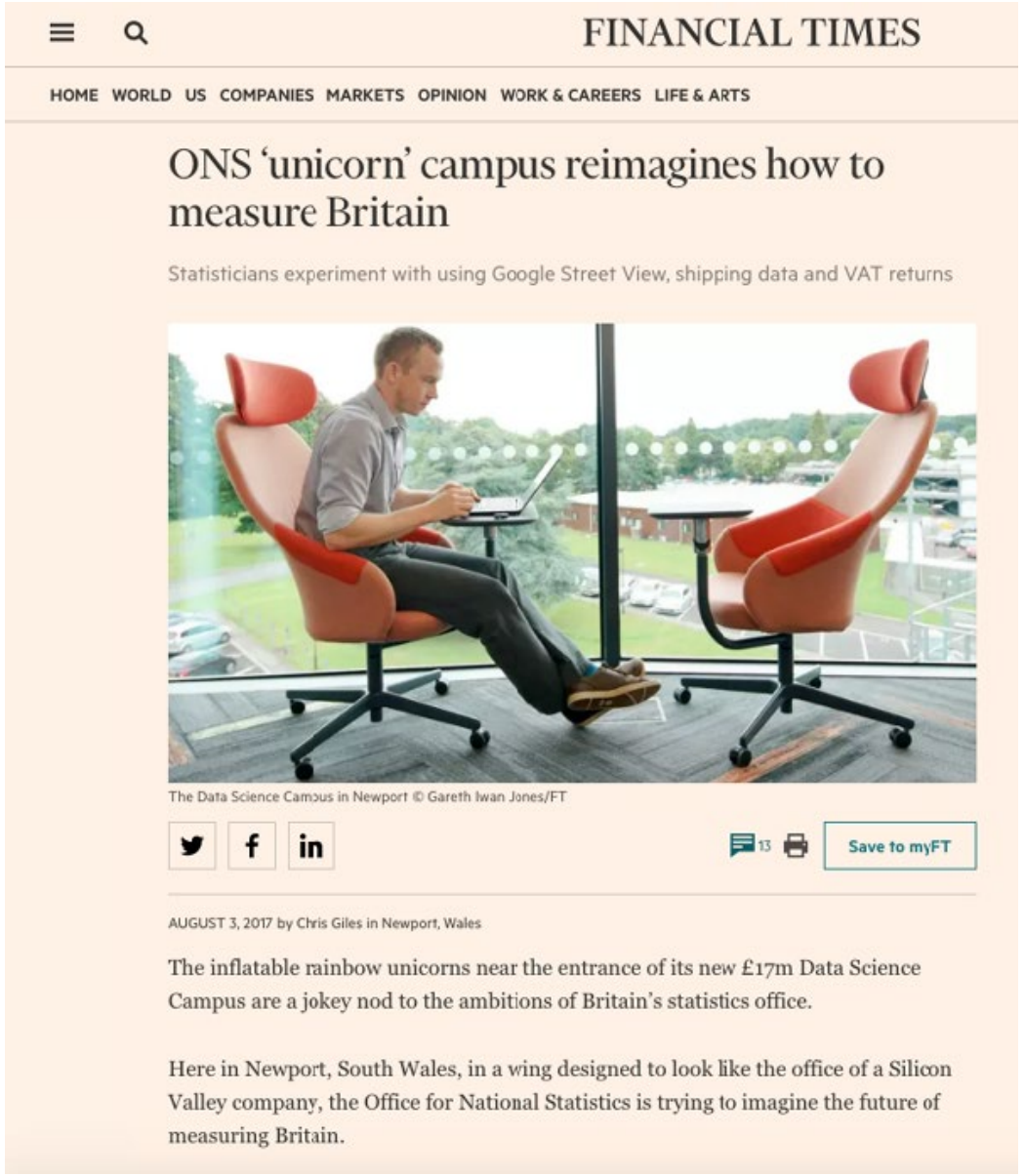
Data Science Campus creation

"Although **better use of [data]** has the potential to transform the provision of economic statistics, ONS will need to **build up its capability** to handle such data.

This will take some time and will require not only **recruitment of a cadre of data scientists** but also **active learning and experimentation**.

That can be facilitated through **collaboration with relevant partners** – in academia, the private and public sectors, and internationally."

Independent Review Economic Statistics
Professor Sir Charles Bean, 2016, p.11



FINANCIAL TIMES

HOME WORLD US COMPANIES MARKETS OPINION WORK & CAREERS LIFE & ARTS

ONS 'unicorn' campus reimagines how to measure Britain

Statisticians experiment with using Google Street View, shipping data and VAT returns

The Data Science Campus in Newport © Gareth Iwan Jones/FT

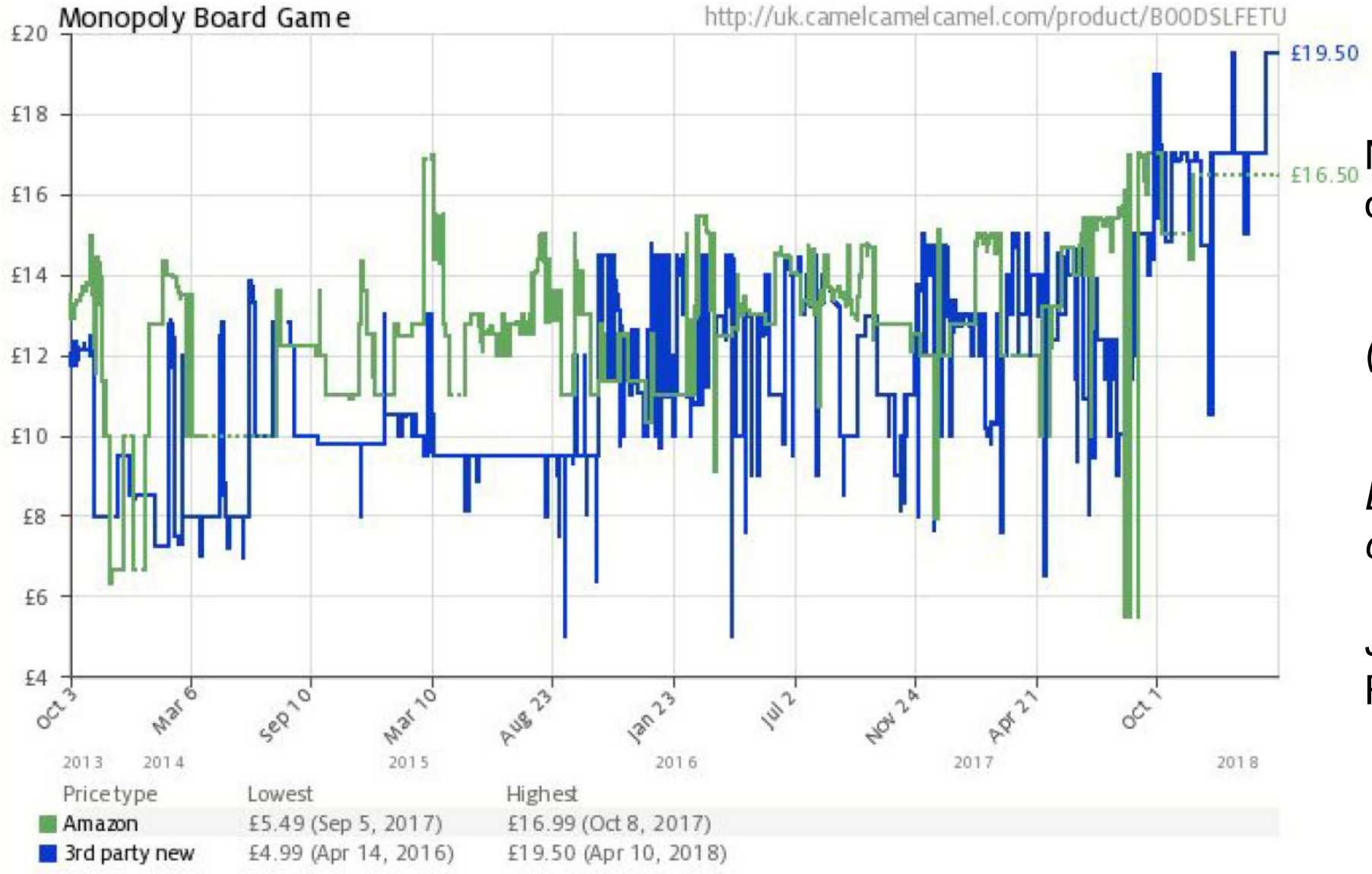
13 Save to myFT

AUGUST 3, 2017 by Chris Giles in Newport, Wales

The inflatable rainbow unicorns near the entrance of its new £17m Data Science Campus are a jokey nod to the ambitions of Britain's statistics office.

Here in Newport, South Wales, in a wing designed to look like the office of a Silicon Valley company, the Office for National Statistics is trying to imagine the future of measuring Britain.

We need big data to understand what is going on!



Monopoly price fluctuation over 4 year period
High = £19.50
Low = £4.99
(Data from camelcamel)

Big Data is changing how consumer markets work

James Plunkett, 2017-18 Rybczynksi Prize Essay

“The 21st Century has brought new challenges in the analysis of data, and it is increasingly apparent that solutions to these are both statistical and computational. This has led to a great demand for people both in industry and in research who are able to draw upon the mathematics of both computation and probability to make sense of the large amounts of data that are collected in order to solve major problems.

Data science is an interdisciplinary response to this demand”

- University of Warwick

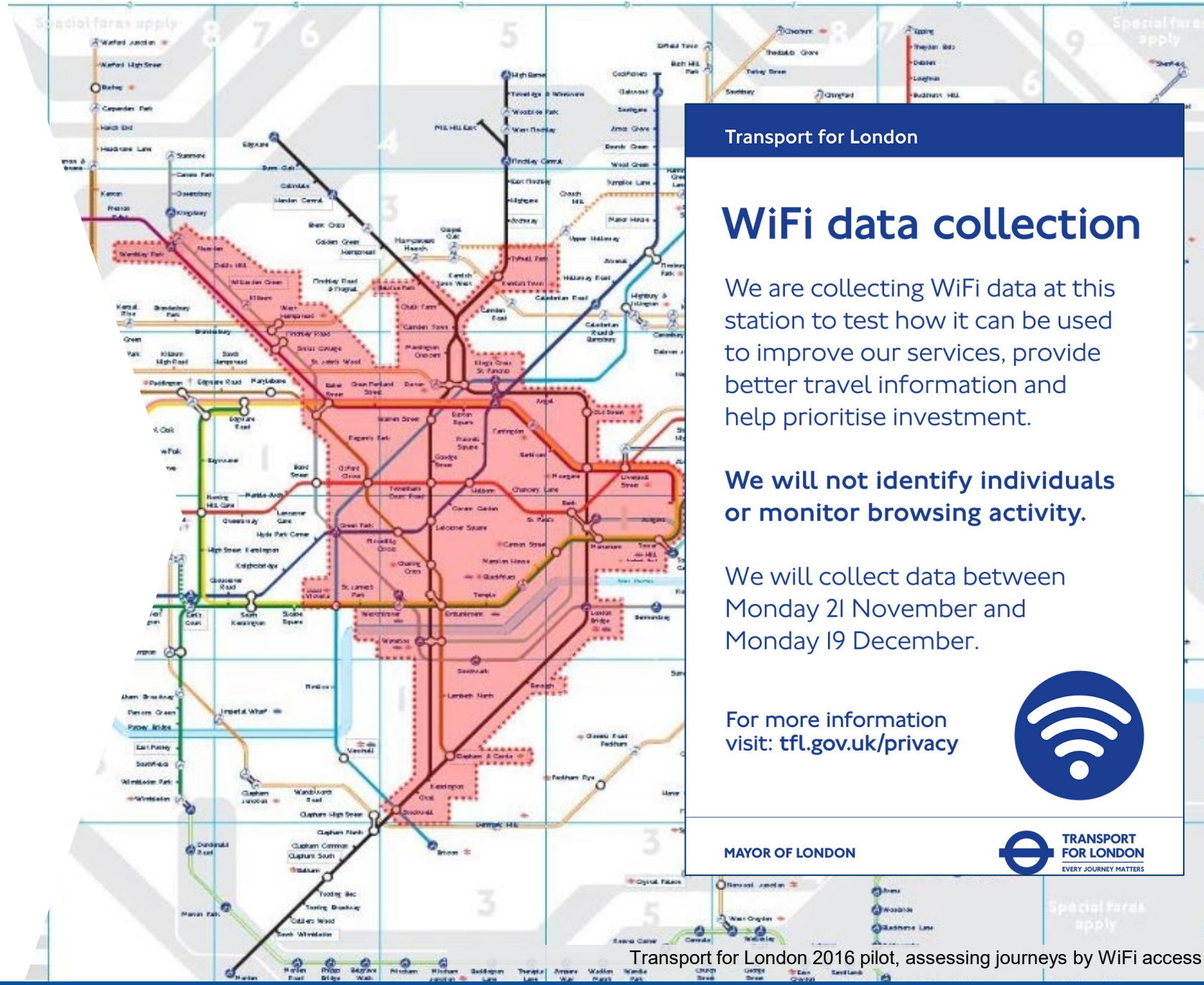
London Transport workers manually examine over 4 million tickets to identify most and least popular routes, March 1939

Gerry Cranham/Fox Photos/Hulton Archive/Getty Images

“The 21st Century has brought new challenges in the analysis of data, and it is increasingly apparent that solutions to these are both statistical and computational. This has led to a great demand for people both in industry and in research who are able to draw upon the mathematics of both computation and probability to make sense of the large amounts of data that are collected in order to solve major problems.

Data science is an interdisciplinary response to this demand”

- University of Warwick



Transport for London 2016 pilot, assessing journeys by WiFi access

Purpose

We apply data science, and build skills, for public good across the UK and internationally

Mission

We work at the frontier of data science and AI - building skills and applying tools, methods and practices - to create new understanding which improves decision-making for public good

Data science for public good – strategic objectives

DSC1	Deliver better statistics, and strengthen evidence for policy-making & public services, by applying data science tools, techniques & practices	HELPFUL
DSC2	Strengthen our ability to understand the economy and society by assessing the value of new data sources and techniques	INNOVATIVE
DSC3	Grow data science capacity, and support the data science community, across ONS, UK public sector, international statistics agencies & wider	CAPABLE
DSC4	Improve UK public sector access to data and data science skills, by working in partnership with academia, industry and civil society	EFFICIENT
DSC5	Maximise the impact of our programme through working openly and supporting reuse of our work	PROFESSIONAL

“Big data”

“Big data” often means
“data produced by someone else”

And there’s lots of it

Early Indicators of GDP

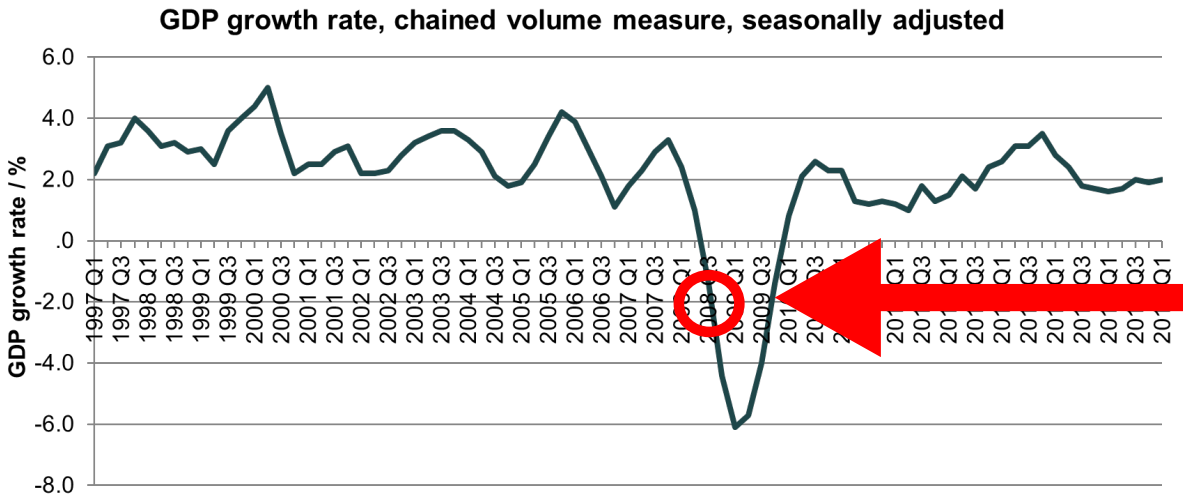


Fig 1. UK GDP Growth Rate

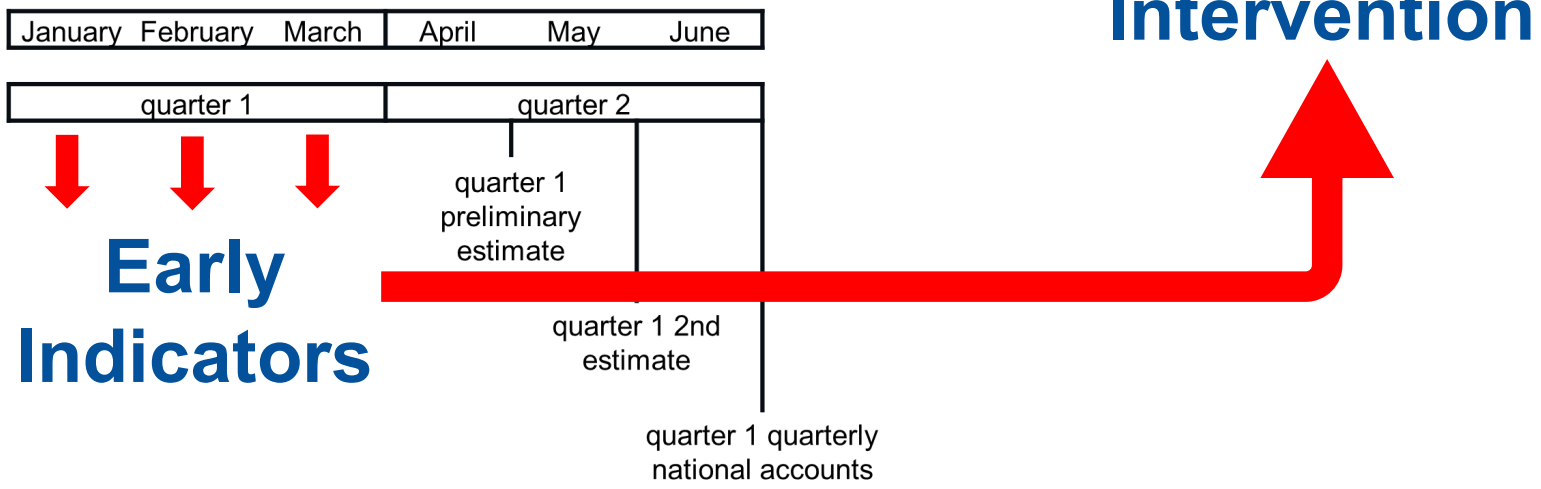


Fig 2. ONS National Accounts Publication Timetable

-6%

Change in UK GDP between first quarter of 2008 and second quarter of 2009

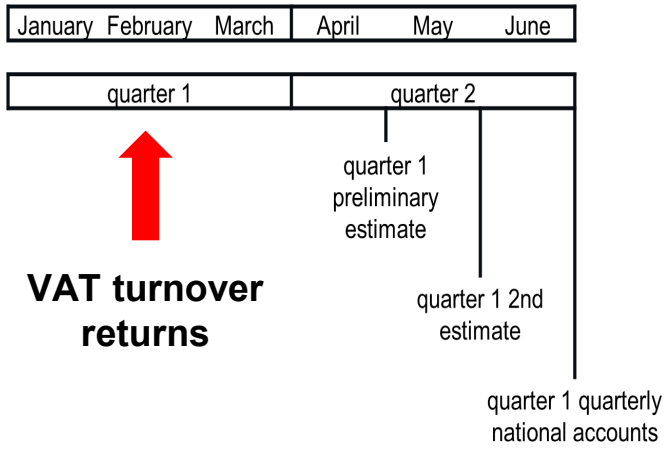
5 years

Length of time from 2008 for the UK economy to return to pre-recession size

£12b

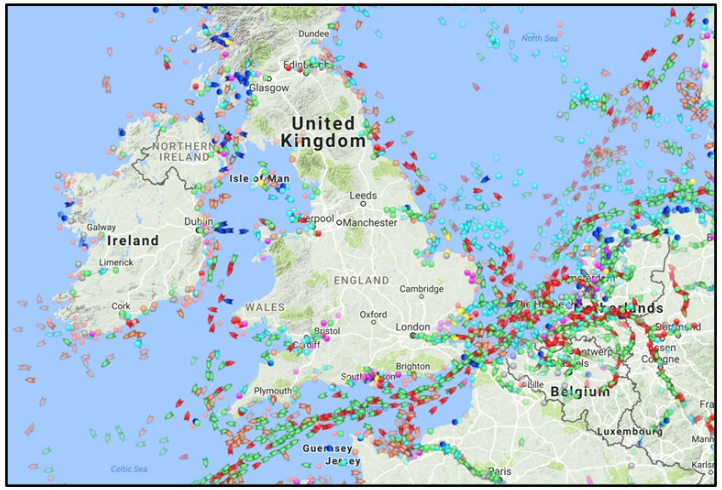
Estimated value for earlier identification of 2008 downturn

Early Indicators of GDP

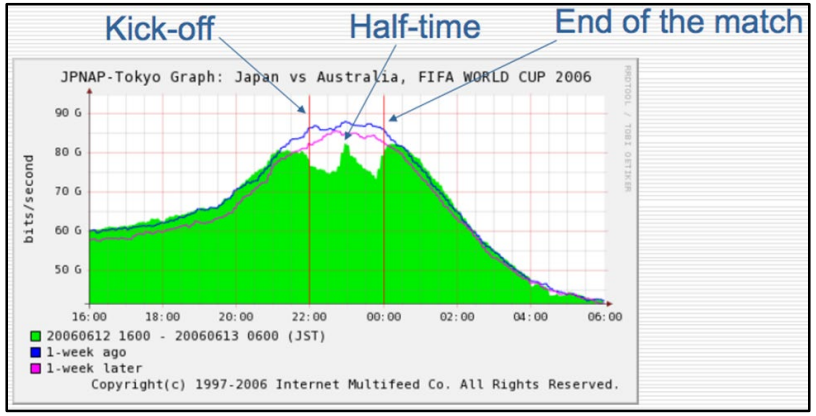


HMRC VAT Data

Road Traffic



AIS Ship Location



Broadband Traffic

-6%

Change in UK GDP between first quarter of 2008 and second quarter of 2009

5 years

Length of time from 2008 for the UK economy to return to pre-recession size

£12b

Estimated value for earlier identification of 2008 downturn

Text analysis of ferry cargo

The Challenge

Ferry operators collect information on the contents of lorries and trade vehicles boarding their Ferries

A single line description is recorded to detail the contents

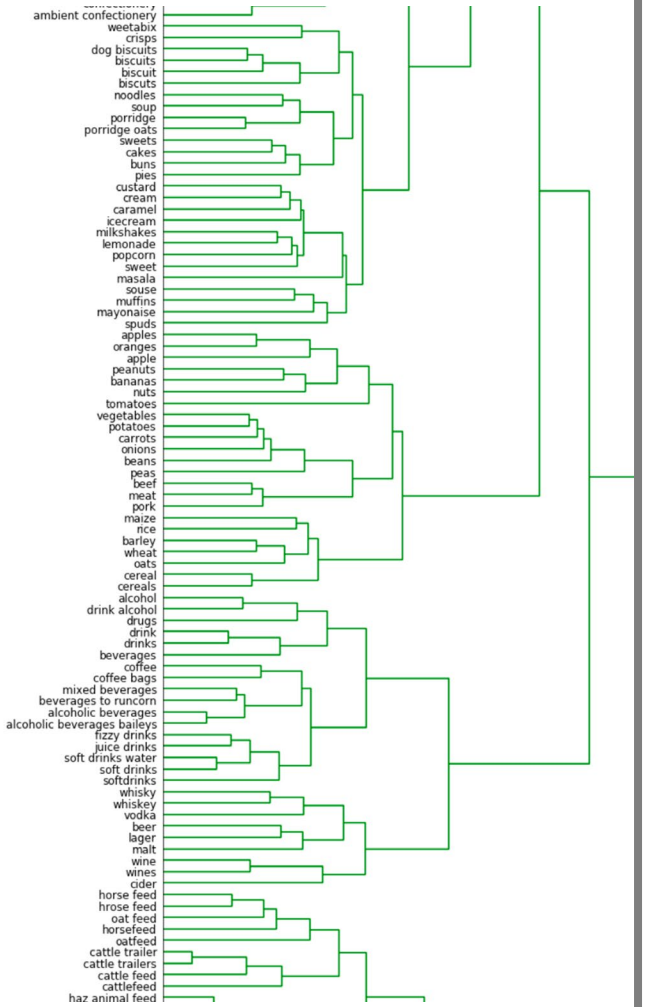
The data collection is not controlled enabling complete free text entries.

This significantly restricts the analysis that can be done.

The Solution

Optimus is an NLP pipeline that can group items from free-text lists by context that do not have accompanying classifications or codes.

The tool can generate labels for groups of items based on common syntax or, in some cases, synonyms. It can also handle inconsistencies in text records such as spelling mistakes, plurality and other syntactic variation.



The Data

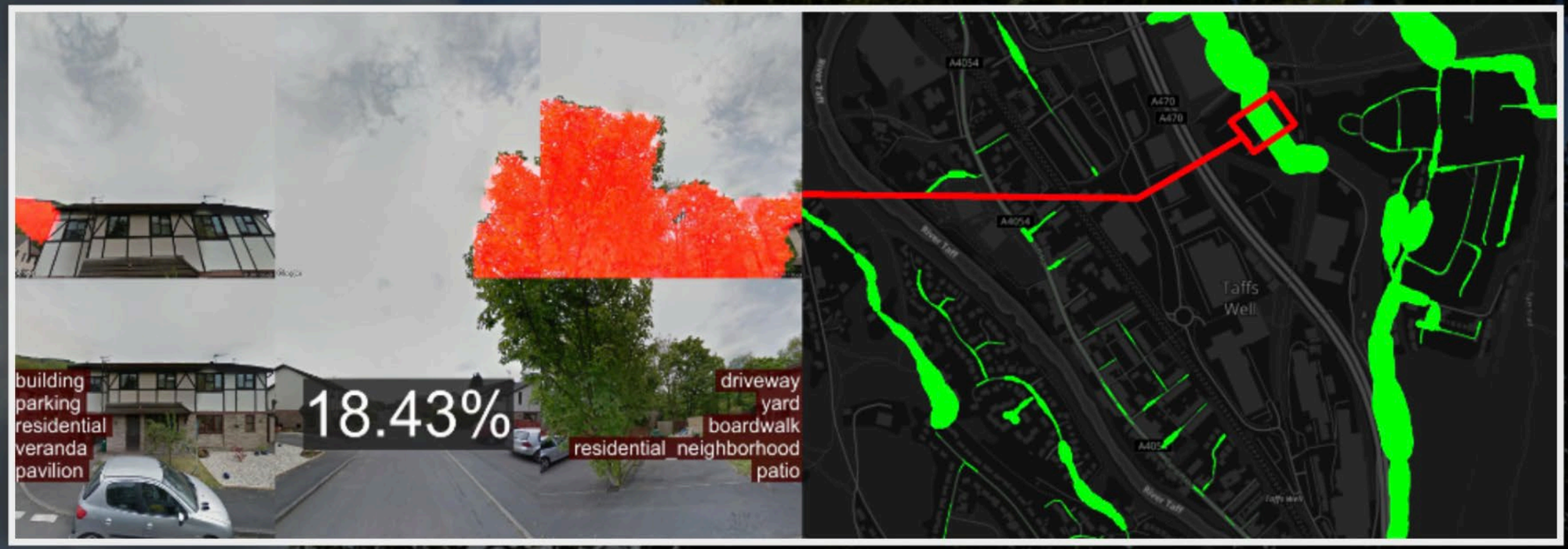
35k

Lorry journeys in single month analysed during Phase 1

450k

Lorry journeys in 2017 to be analysed during Phase 2

Mapping the urban forest



Makes use of:

1. Google streetview imagery
2. OpenStreetMap road network data

Growing skills

Degree level Apprenticeships in Data Analytics:

School leavers plus. 12 months at the Campus followed by 6 month rotations across ONS

Data Science Accelerator:

12-week mentoring programme for Government analysts

Data Science Faculty:

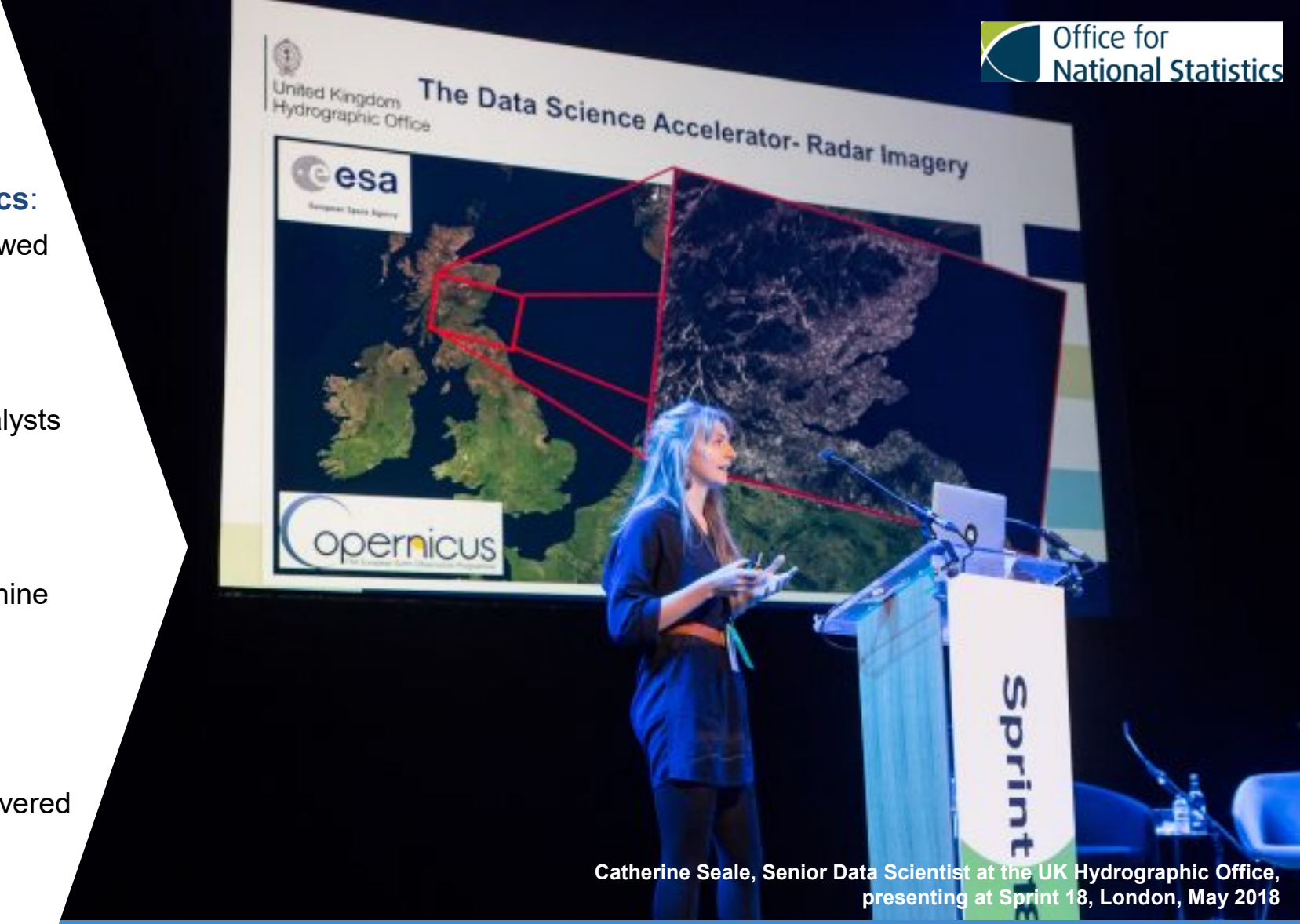
In-house training unit delivering short courses in programming (R, Python) and fundamentals of Machine Learning, NLP, etc. “Art of the Possible” course

Masters in Data Analytics for Government:

Two-year, part-time MSC for government analysts; Continuous Professional Development modules delivered locally in Data Science Faculty

PhD internships:

Part-sponsorship; 3-6 month internships in Campus



Catherine Seale, Senior Data Scientist at the UK Hydrographic Office, presenting at Sprint 18, London, May 2018

Office for National Statistics

Growing Data Science skills across the public sector

Department for International Development

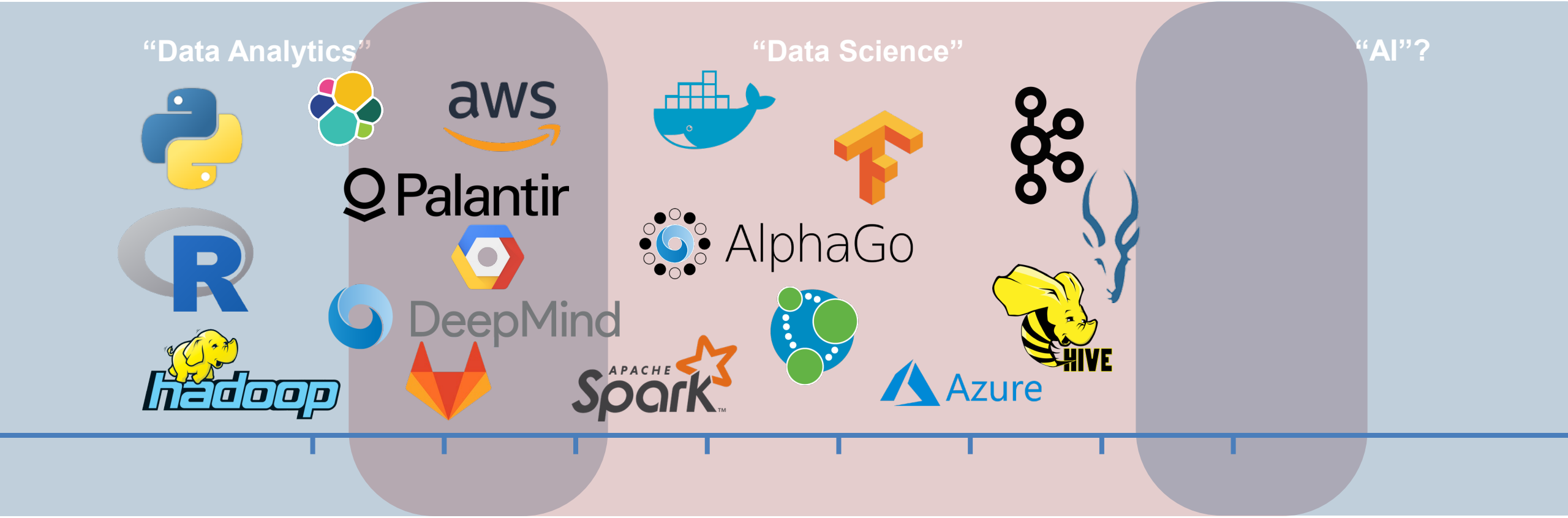
Data Science in the UK Government

A recent history and what we learned from it

Tom Wilkinson
Head of MI and Analytics

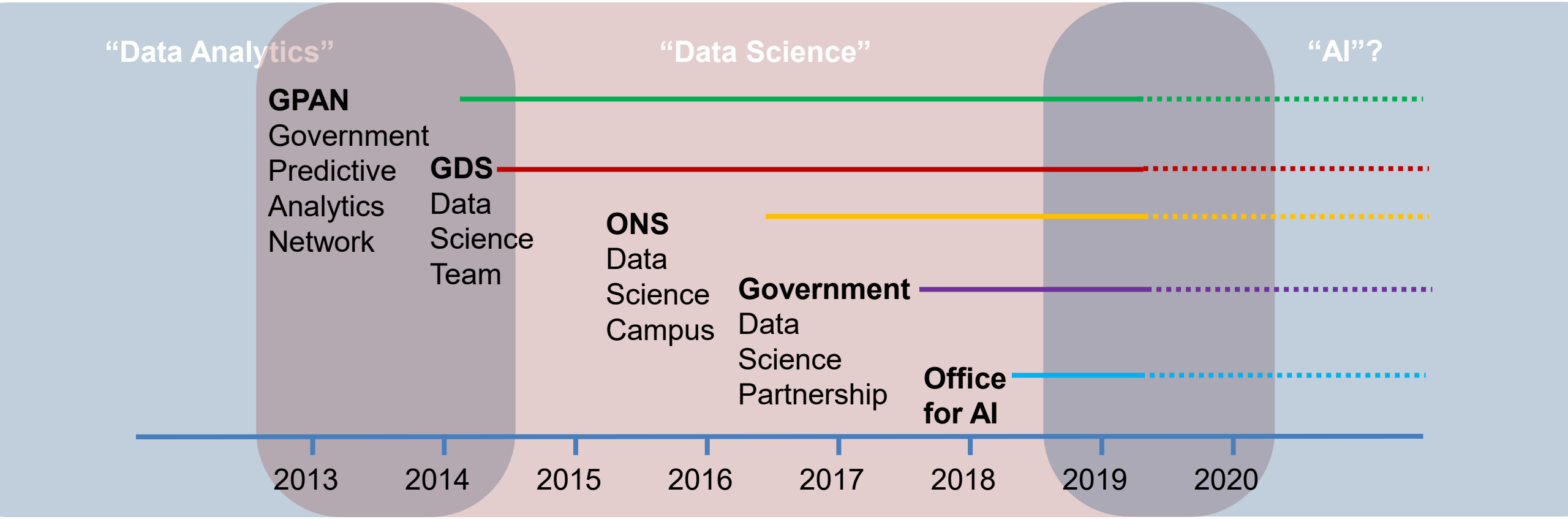
“Data Science” has evolved **continuously**

Department for International Development



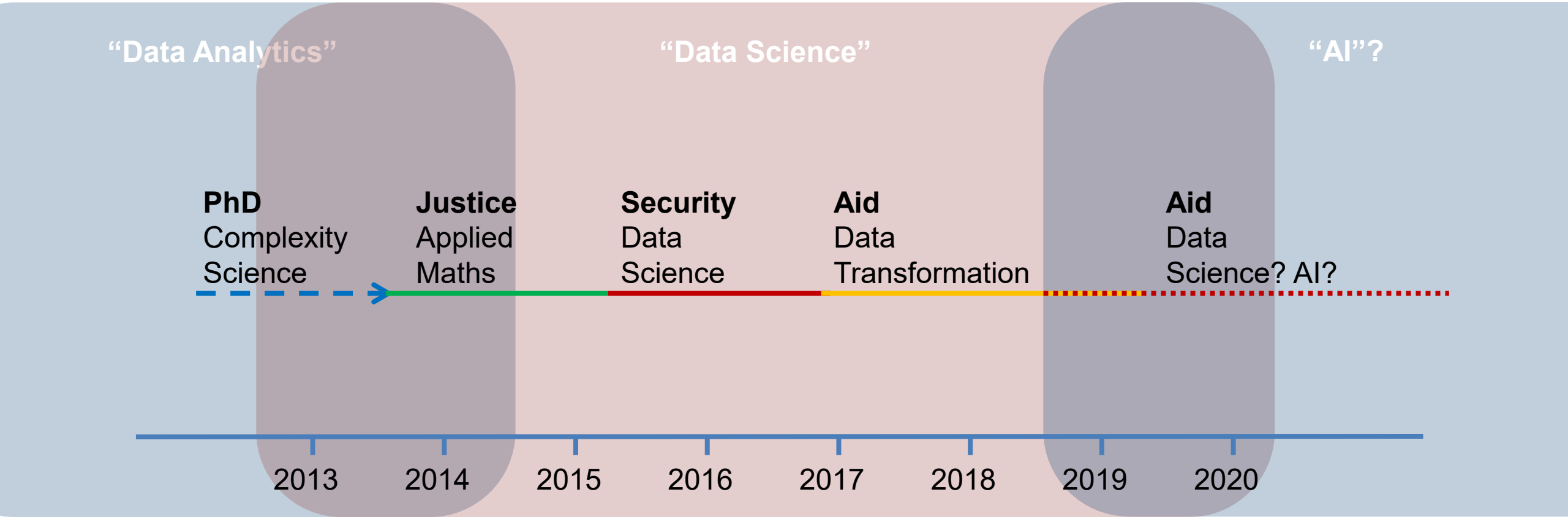
Data Science **in government** has evolved continuously

Department for International Development



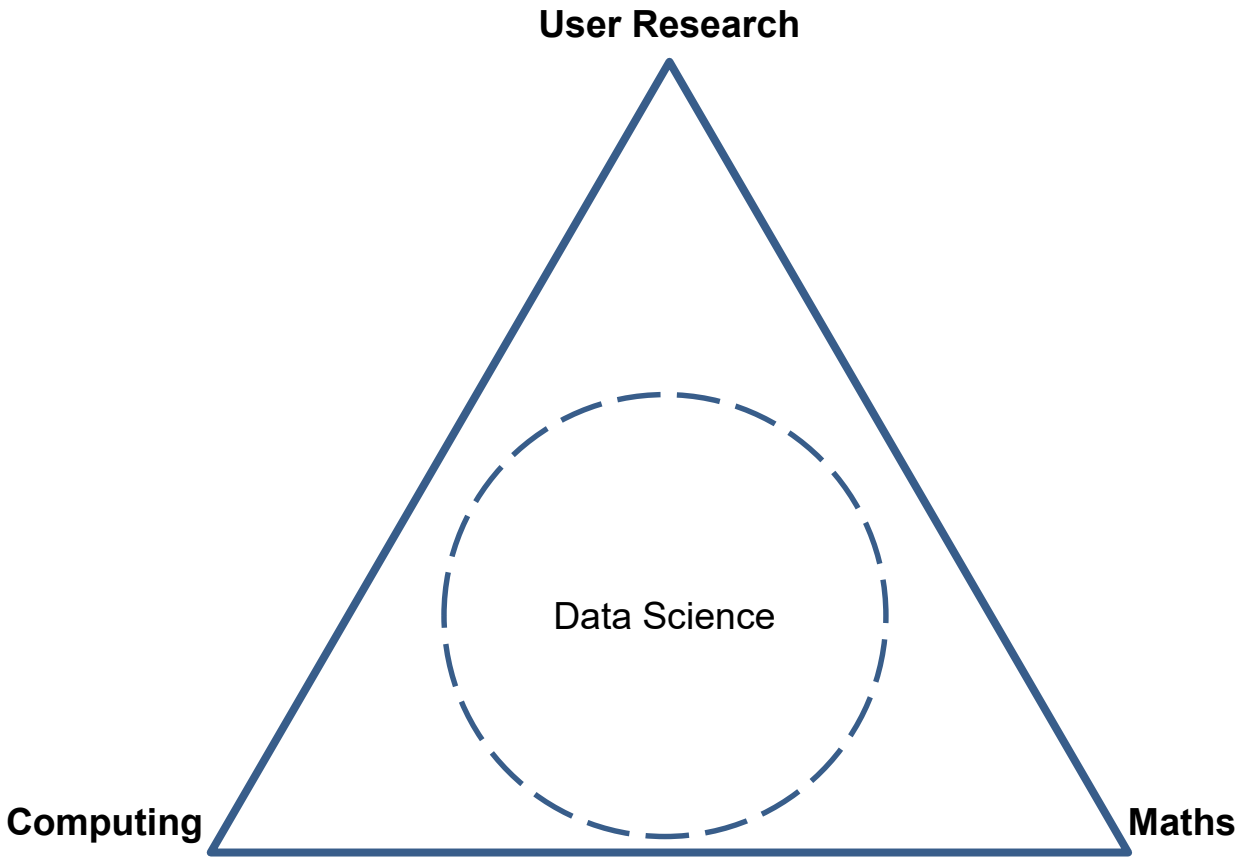
(I've toured various roles over this time)

Department for International Development



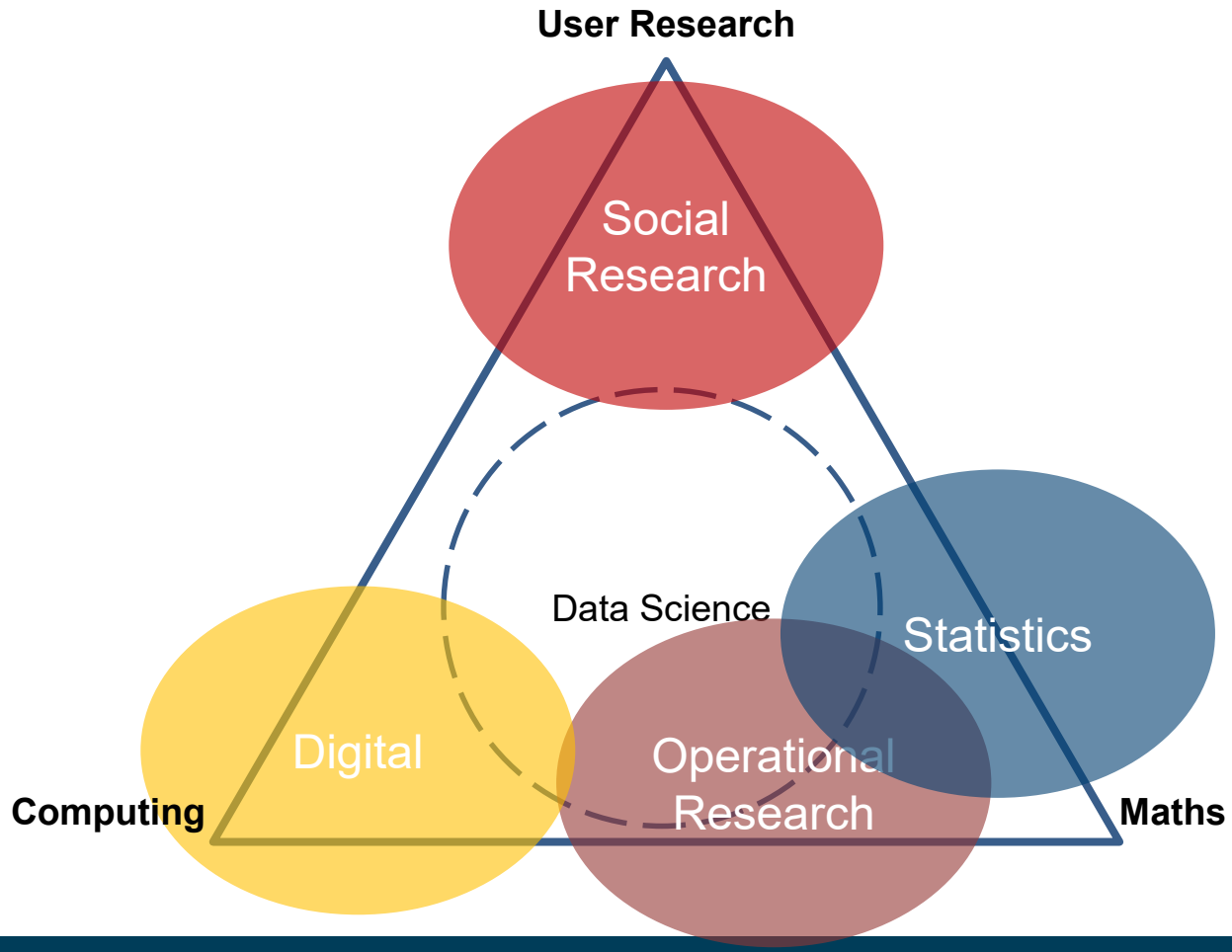
Many groups have pulled together… (mostly)

Department for International Development



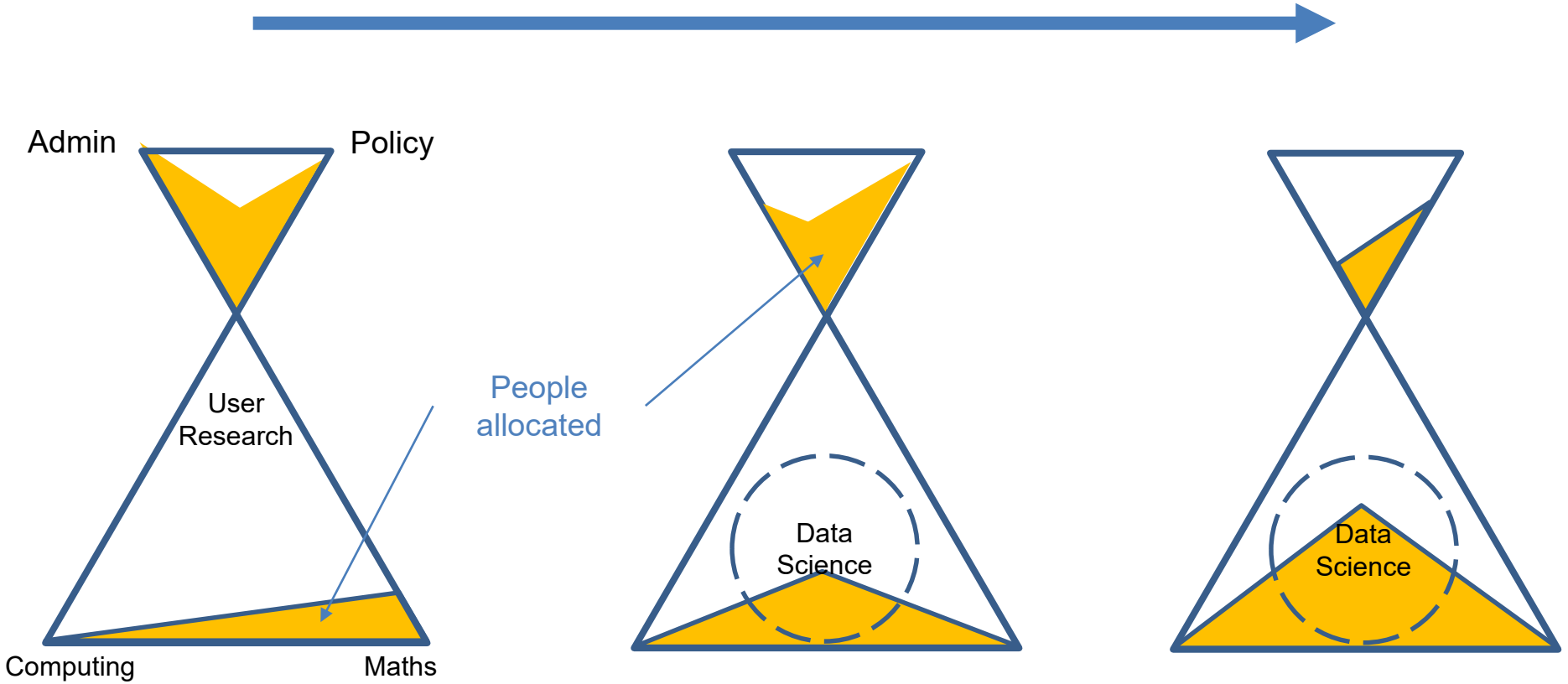
Many groups have pulled together… (mostly)

Department for International Development



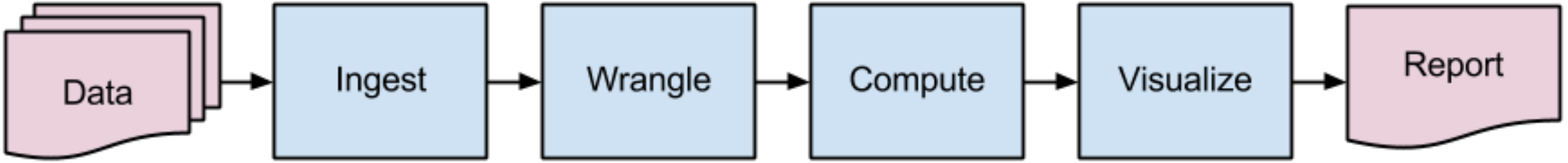
We’re on a good skills trajectory, but we have a way to go

Department for International Development



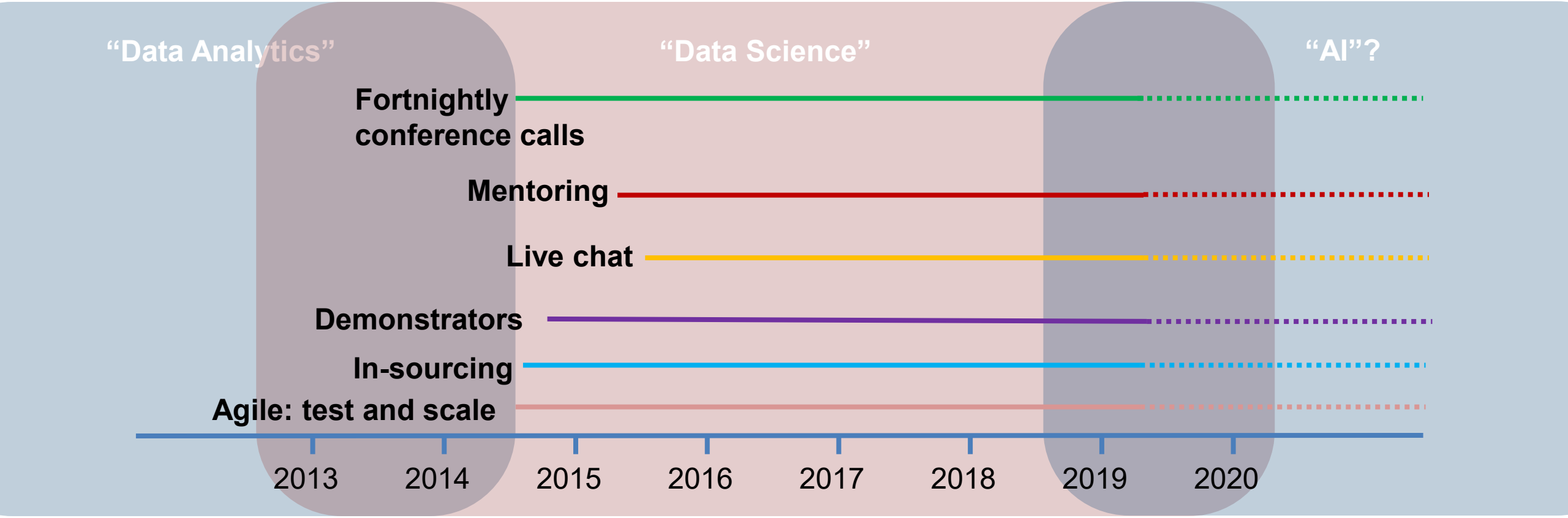
Data infrastructure and sharing is equally important

Department for International Development



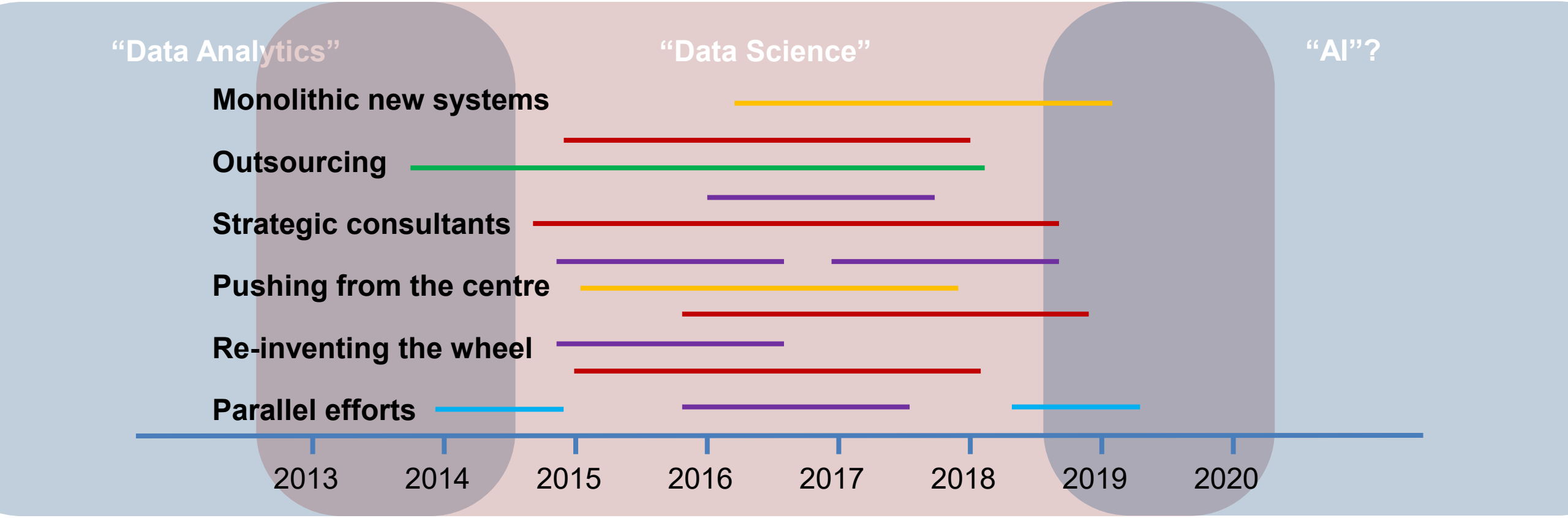
Agile, bottom-up, collaborations have worked well

Department for International Development



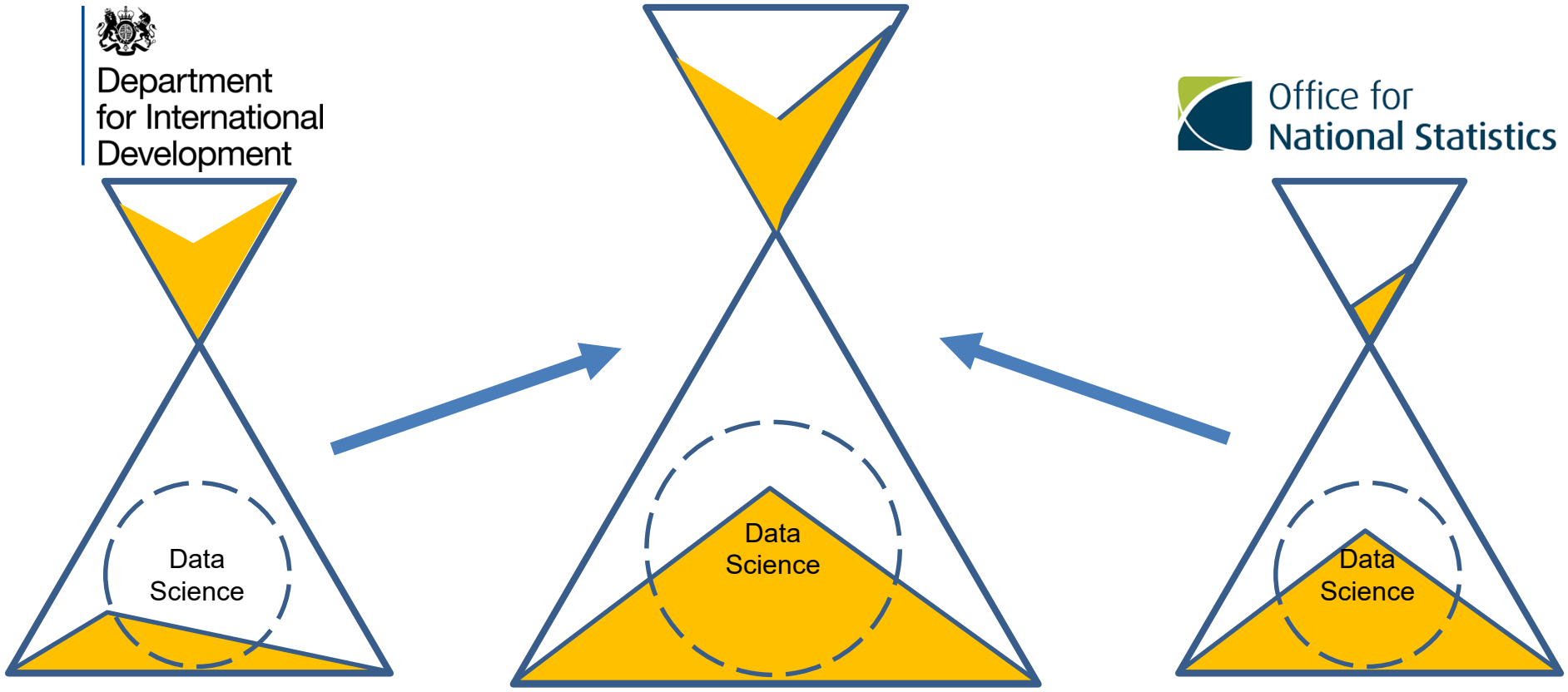
Top down, outsourcing, and disconnected parallel work hasn't

Department for International Development



Applying this learning by partnering ONS technical expertise with DFID's aid experience

Department for International Development



Building Data Science Capacity at ONS and Beyond

Monday 29th April, 2019

Ceri Regan
Academic Manager
UK Data Science Campus

Overview

- Building Data Science Capability in a Government Department or NSI
- How we are doing this at the UK Data Science Campus
- Inspiring a culture of innovation
- Building other data science capacity
- Our Partnership with NISR
- Our work with UNECA
- Closing remarks – Tom Smith
- Discussion

Building Data Science Capability

Three main routes followed by UK Data Science Campus

Influence future recruitment

Knowledge Exchange

Recruitment

- Grade structure required for the team?
- Qualifications/experience required at each level?
- Size of pot to recruit from?
- Direct/target your recruitment activities
- Consider the apprenticeship route?

Grow your own

- Build on analytical skills
- Not all statisticians need to become data scientists
- Offer choice of data science training, including qualifications
- Direct Gov collaborations
- Encourage sharing of experiences/knowledge across teams

Draw on skills from elsewhere

Academia

- Joint research programmes
- MSc & PhD placement students / theses
- Academic secondments

Industry

- Public good outputs with key partners
- Secondments
- Knowledge sharing events/ hackathons
- Data sharing

Building Data Science Capability

Vocational Apprenticeships

- Work and study for a BSc in Data Science
- On the job training, week release to University
  - Salary is paid by NSI
  - University fees paid by Gov
- Future – MSc in Data Science Apprenticeship

Operational – Campus Faculty

- Self sufficiency – develop Champions
- Develop & deliver curriculum: R, Python, Spark, NLP - knowledge exchange
- Manage and administer:
  - Fortnightly seminars
  - 12 week Accelerator programme
- Provide consultancy – offer capability building solutions

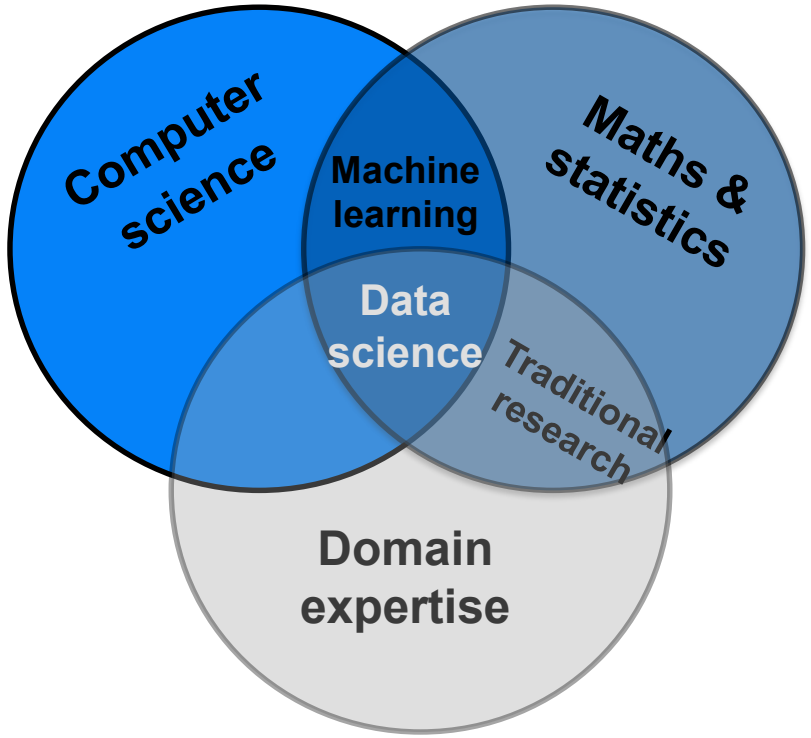
Academic

- MSc in Data Analytics for Government – part-time
  - Southampton University
  - University College London
  - Oxford Brookes University
  - Others joining soon…Glasgow (online MSc), Cardiff, etc
  - We offer funding for 10 UK gov staff per annum
- MSc/PhD placement students – undertake government projects for thesis

Building Data Science Capability

Building on Analytical Skills across Gov

- Leading the development of data science skills
- Supporting and upskilling Gov Analysts
- Understanding current skillset
- Building a picture of learning gaps
- Developing career pathway
- Developing L&D pathway/curriculum



The Data Science Venn Diagram, designed by Drew Conway

Inspiring an Innovative Culture

Inspiring Senior Managers

- They are the catalyst
- Showcase the DS work taking place
- Show what is possible - **inspire**
- Hold discussions around ‘barriers’ to innovation

Ensure all staff are ‘aware’

- What is Data Science/Big Data/Artificial Intelligence?
- What does this mean for me?
- What does it mean for the dept/NSI?
- Why are we doing this?
- Show what is possible - **inspire**

Building Data Science Capacity

- It's not just about the programming/Machine Learning/NLP skills

To build **NSI capacity**, you may also need to consider:

- IT infrastructure – for storing and analysing data
- The right landscape – legal frameworks, data access
- Ethics – ensure public trust

We need to draw on other ONS experts to assist us

Partnership with NIS Rwanda

- Through DFID partnership
- Data Revolution in Rwanda
- We have provided Consultancy:
  - Building out Data Science research teams
  - Building Data Science Capability
  - IT infrastructure
  - Legal framework/data access

Alex Noyvirt delivering a Python workshop at NISR, Kigali, October 2018

- Current status:
  - Legal and IT discussions continue
  - Established two joint projects with the UK Data Science Campus
  - UK is providing mentorship and training
  - Aiming for self sufficiency

ONS working with UNECA

- Through the DFID partnership
- ONS and Data Science Campus are working with UNECA in various ways:
  - Consulting on design of the Campus
  - Advising on SDG data gaps
  - Census/data quality training
  - Preparing to deliver Python/NLP training
  - Planning joint projects
- Further Consultancy to establish other learning needs

Team from ECASTATS discussing SDGs with ONS and Data Science Campus, December 2018

Closing Remarks

- Not every country will need to develop capacity at the level of UK/Rwanda/UNECA
- Different models exist – it's finding what works for you and your needs
- Working in partnership with others can have a large impact
- We are all trying to learn what works – ideas and experience are welcome

Discussion

- What are the data science skills needs in your NSO?
- How are these skills needs already being met?
- What more can be done to develop these?