Data Science Campus

Growing public sector data science capability

Data4Good conference
16th October 2020

Tom Smith, Director
ONS Data Science Campus
@_datasmith

The Role of the Campus

Evidence produced free from political and commercial influence is a cornerstone of democratic society. It allows for the better formulation and evaluation of public policy; it informs public understanding of the world and our place in it; it informs decision makers to take choices and lets others hold the powerful to account.

The Authority is uniquely placed for its role in this through...

> **5. the capacity to innovate, develop new data science techniques and respond in real-time to the nation's needs through the Data Science Campus;** and
>
> 6. the practical application of data ethics, through the National Statistician's Data Ethics Committee and **the exercise of impartiality, transparency and openness in all we do to retain the trust of society**.

Statistics for the Public Good, UKSA, p.11

International Programmes

Successful collaboration in the modern statistical world must be international too. As the pandemic and environmental issues have vividly demonstrated, users need data that can be understood in the international context, not in national isolation. **We will provide global statistical leadership** in those fora, and on subjects central to the UK's interests.

We will promote transparency and high standards, and support coherence and comparability across the global statistical community. And **we will continue to work with, and learn from, our partners in the developing world to ensure statistics and data are at the heart of the UK's development policies and agenda, and provide help to those most in need**.

Statistics for the Public Good, UKSA, p.18

“Data … has been a lifeline during the global coronavirus pandemic

The fact that governments, businesses, organisations and public services were able to share vital information quickly, efficiently and ethically during the pandemic has not only saved lives but has enabled us to work from home, keep the economy running and stay connected with loved ones during a period of unprecedented disruption. As we enter into recovery, it is vital that we make the most of what we have learnt.

[This] means driving **a radical transformation of how the government understands and unlocks the value of its own data to improve a range of public services and inform decisions at scale**, through a whole-government approach driven from the centre.”

Rt Hon. Oliver Dowden CBE MP (Secretary of State for Digital, Culture, Media and Sport)
Ministerial Forward to the National Data Strategy

Policy paper

National Data Strategy

Published 9 September 2020

With further measures to be announced as part of the Digital Strategy and through the National Skills Fund, we will:

- prioritise bringing in and building the right skills across government
- recruit leaders with data and digital skills across government to build a strong cadre of technical, policy, legal and analytical data experts in the centre of government
- train 500 analysts across the public sector in data science by 2021, through the Data Science Campus at the ONS, the Government Analysis Function and the Government Digital Service. This will be reviewed in 2021 with a new capacity building strategy meeting the emerging needs of government up to 2025.
- deliver the range of actions to be outlined in the Public Sector Data Science Capability Audit
- review data training available to all civil servants and develop proposals to enhance and extend this offering
- design a career pathway for data expertise in government
- agree a shared definition of data expertise across central government
- review the needs of local government in having the capabilities to manage, use and disseminate data

Public Sector Data Science Audit

Government Data Science Partnership

The Public Sector Data Science Capability Audit undertaken by the Campus and the Government Digital Service (GDS) through the Government Data Science Partnership (GDSP) found that to advance data science across the public sector, organisations need to:

1) understand the contribution that data science can make to strategic objectives;
2) develop a culture of experimentation; and
3) ensure that their data science staff have access to the data, tools and technology they need for effective analysis.

The Audit identified a range of Areas for Action under six key themes:

Data; Technology; People; Application; Ethics/Governance; Organisation Culture

Capability Programmes

Data Science capacity-building for the Public Sector

Data Science Faculty

- In-house training unit delivering operational courses focused on up-skilling government analysts in data science skills, from fundamentals of computer science and programming (R, Python) to data science (Machine Learning, NLP etc) and data engineering.
- Data Science Lecturers recruited from universities to develop and deliver a core AI and data science curriculum for use across government.
- Directly supports internal and external partners at an organisational level through bespoke capacity-building programmes.

Mentoring Programmes

- Accelerator programme: 12-week cross-government mentoring programme for Public Sector analysts delivered with GDS, run in parallel with in-house ONS Data Science Academy.
- Graduate Data Scientist Programme: 2-year programme for PhD and MSc with second cohort starting in October 2020.

International Programmes

- Data Science for Official Statistics: Faculty-delivered workshops, training and mentoring programmes for National Statistics Organisations
- Data Science for International Development: Dedicated research and training team co-located with Department for International Development in Glasgow

Cabinet Office | Companies House | Department for Business, Energy & Industrial Strategy | Department for International Development | Department for International Trade | Department for Work & Pensions | Home Office | Ministry of Justice | NHS Digital | Ofcom making communications work for everyone

Office for National Statistics | Northern Ireland Statistics and Research Agency Gníomhaireacht Thuaisceart Éireann um Staitisticí agus Taighde | Government Statistical Service | NISR National Institute of Statistics of Rwanda | United Nations Economic Commission for Africa

Academic Programmes

Developing the next generation of government data scientists

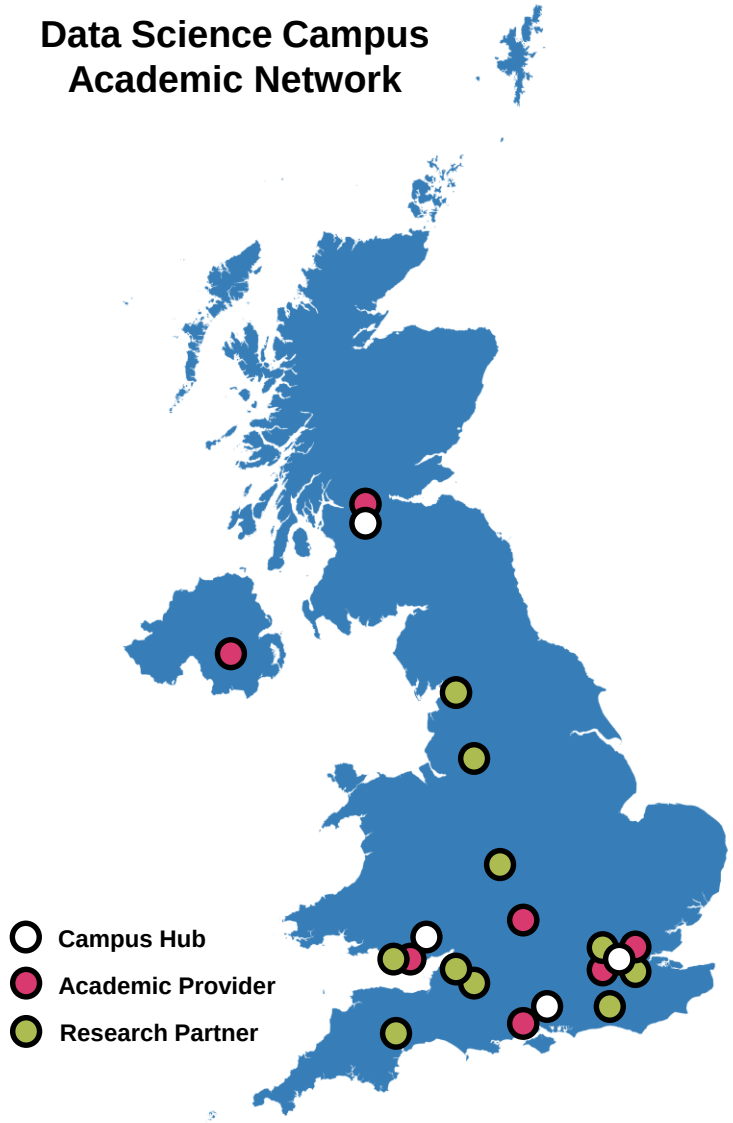
Research Partnerships

- The Data Science Campus works with 7 UKRI, EPSRC and STFC Centres for Doctoral Training in AI, Data Science, Statistics and Physics, with collaborative research, PhD mentoring and three-month industry placements at the Campus for both PhD and MSc students.
- A collaborative partnership with the Alan Turing Institute, focused primarily on economic data science, and a *Data Science for Public Good* joint PhD programme.
- Direct partnerships with Cardiff University, Lancaster University, University of Manchester and University of Warwick that facilitate collaborative research, PhD co-funding, MSc and PhD placements and academic staff secondments.

Academic Partnerships

- Level 6 Apprenticeship in Data Science: Framework development in England and Wales led by the Campus, The first three-year programme launched in Wales in 2019 with Cardiff Metropolitan University building on 2 successful Campus cohorts of Level 4 Apprenticeship in Data Analytics.
- Masters in Data Analytics for Government: 2 year part-time MSc in Data Science and Statistics for government analysts, delivered by Cardiff University, University of Glasgow, Oxford Brookes, University of Southampton and UCL.

Data Science Campus Academic Network



Government Data Science Partnership

Government Data Science Partnership

Championing Data Science in government

- Established in 2015 as a collaboration between the Government Digital Service, the Office for National Statistics and the Government Office for Science, to support the growth of data science across government and the developing community of practitioners.

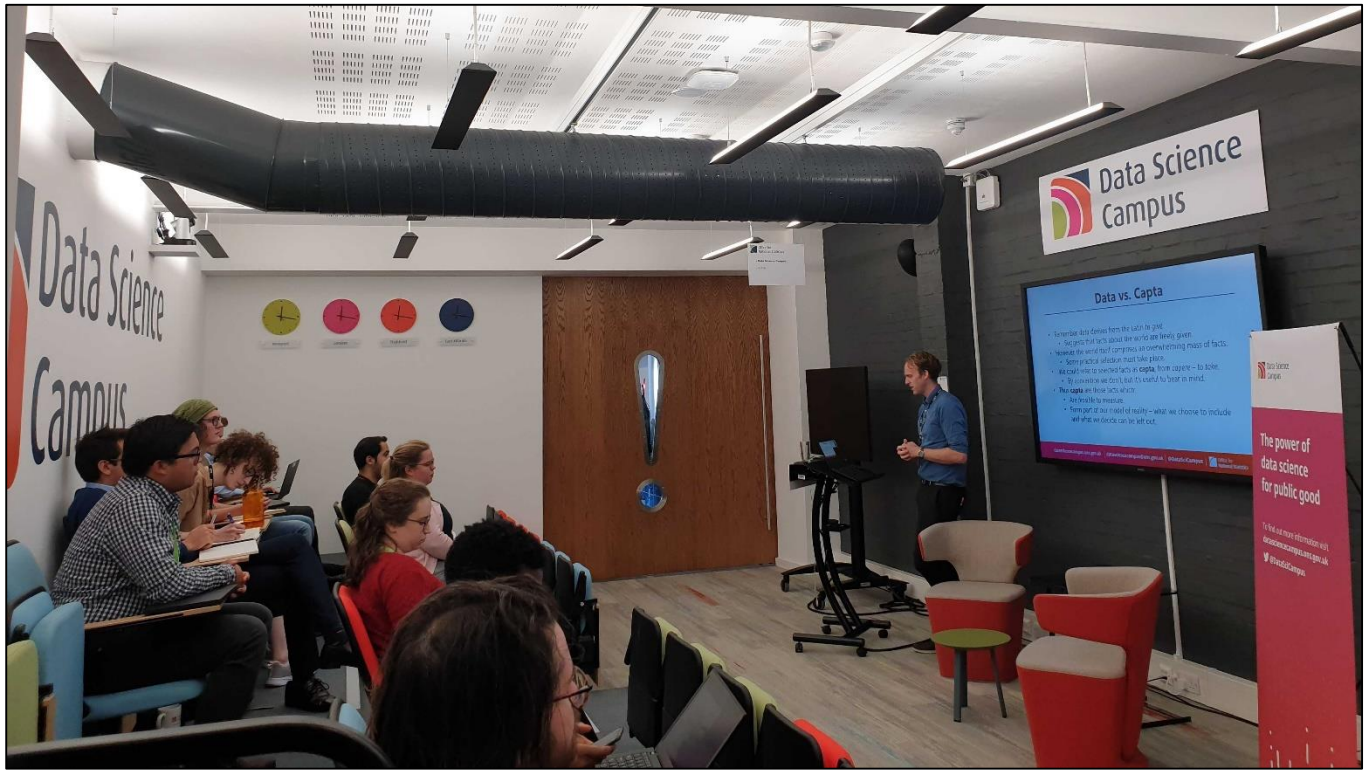
Building Skills and Standards

- Accelerator programme: 12-week cross-government mentoring programme for Public Sector analysts. 16 cohorts to date with 7 mentoring hubs in London (GDS and ONS), Manchester (HMRC), Newcastle (DWP) Newport (ONS), Sheffield (DfE), Taunton (UKHO).
- Public Sector Data Science Audit: An audit of public sector data science capability, covering national, devolved and local government, health services and police. To be published in 2020.
- Skills Working Group: A cross-government working group agreeing standards in data science skills and roles, it launched a new government Data Scientist job profile in Feb 2020.

Building Community

- Two full-time community managers (1 x GDS, 1 x ONS).
- Gov Data Science Slack: An online community of interest of over 2,800 members.
- Government Data Science Conference: Now in its third year, this annual event hosted nearly 300 data science practitioners from across the UK public sector in Manchester in December.
- Community of Interest events: Held quarterly, these knowledge-sharing events bring together up data science practitioners from across the public sector and in the last year have been held in London, Manchester, Southampton and online.

Graduate Data Scientist Programme



The Graduate Data Scientist Programme started 2019 with the first Cohort of HEO Graduate Data Scientists recruited from recent MSc and PhD graduates.

This is an ONS-wide programme run by the Campus, with 6 graduates now based at the Campus - 3 in London, 2 in Newport and 1 in East Kilbride - and 2 based in the Digital Services and Technology Directorate in Newport.

The two-year programme consists of a twelve-month programme of advanced training delivered in-house by Campus, combined with three six-month research placements and a six-month teaching placement in ONS.

Courses are also open to other participants across ONS, and this programme is intended to be a **pilot for a government-wide graduate training programme**.

Faster indicators of the economy

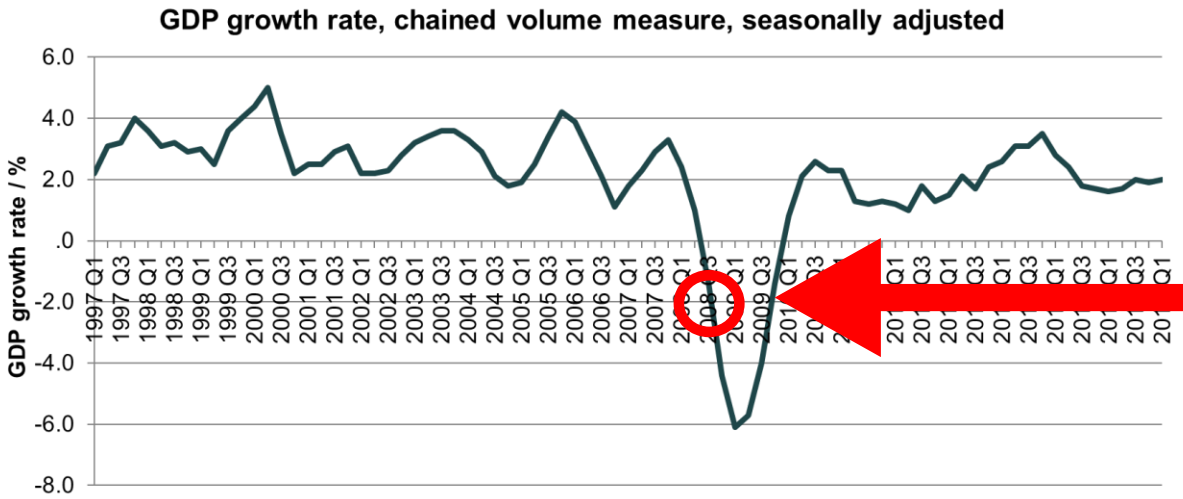


Fig 1. UK GDP Growth Rate

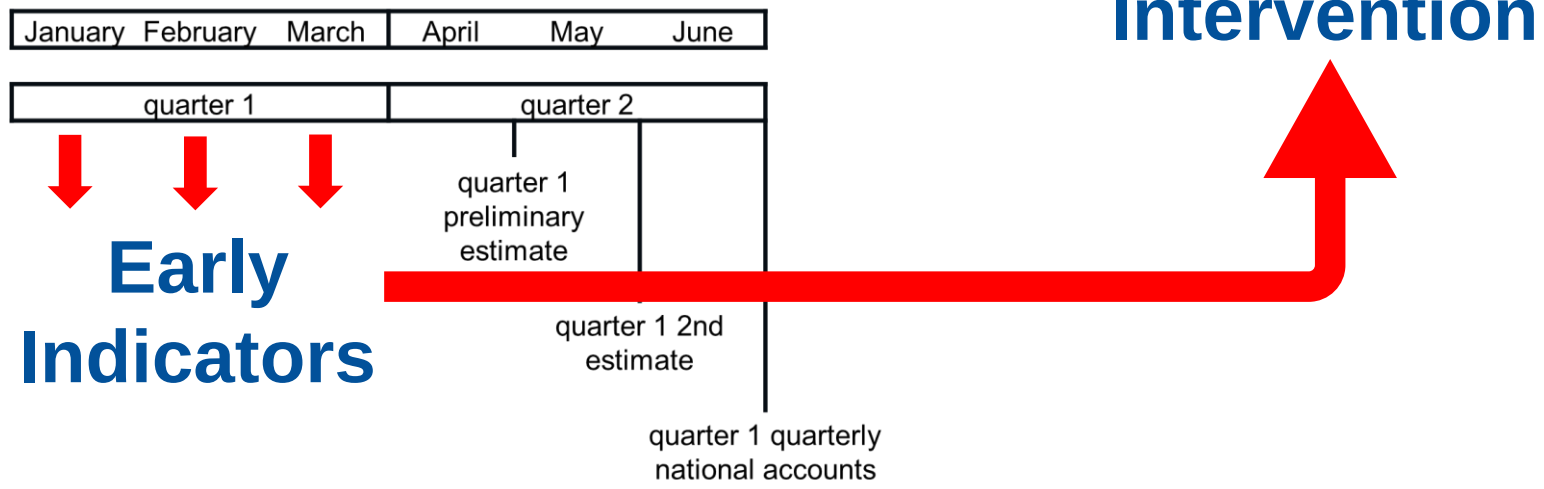


Fig 2. ONS National Accounts Publication Timetable

-6%

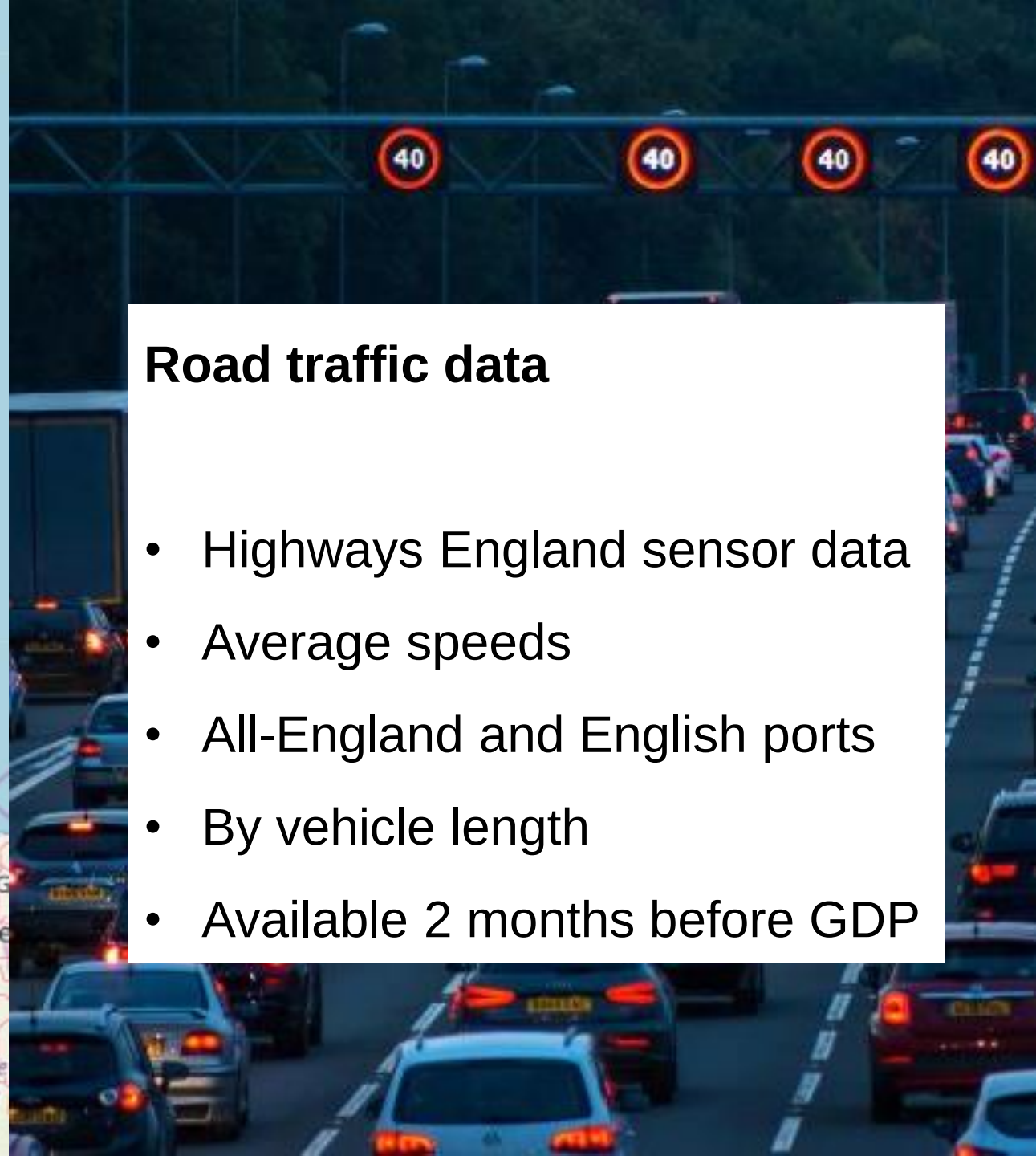
Change in UK GDP between first quarter of 2008 and second quarter of 2009

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

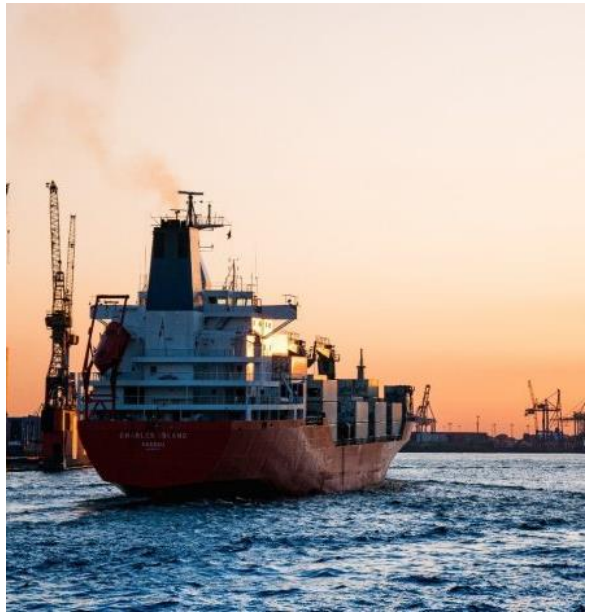


Road traffic data

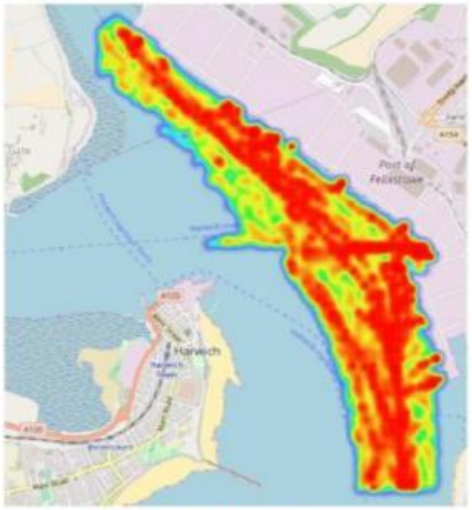
- Highways England sensor data
- Average speeds
- All-England and English ports
- By vehicle length
- Available 2 months before GDP

Shipping Automated Information System (AIS) data

- Marine and Coastguard Agency, ORBCOMM, Global Platform
- Ship tracking data
- Port traffic frequency
- Time in port
- Real time



Monday 12 December 2016



Monday 19 December 2016

Sunday 25 December 2016

Sunday 1 January 2016

Text analysis of ferry cargo - Optimus

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

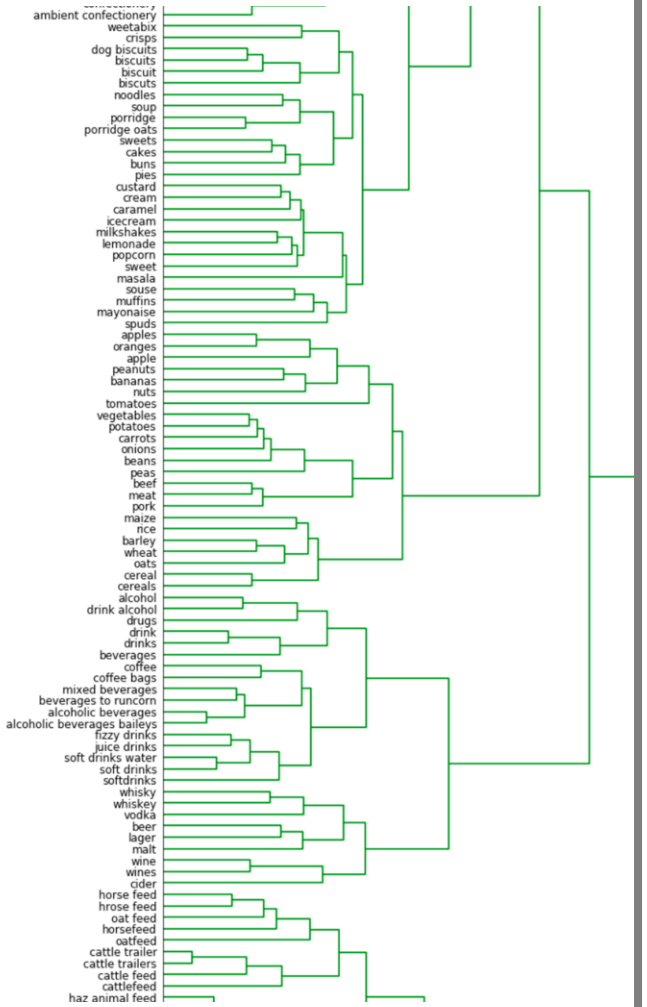
The data collection is not controlled enabling complete free text entries.

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

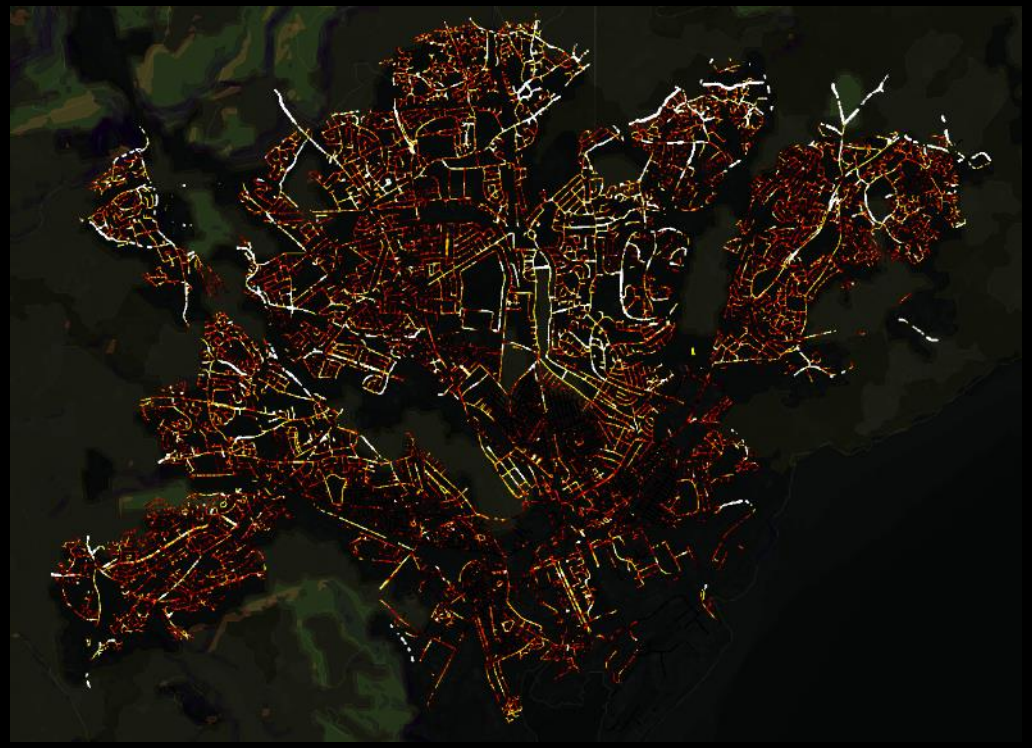
Lorry journeys analysed during Phase 2

Aim: Generate a scalable, consistent, automated, **urban vegetation index**

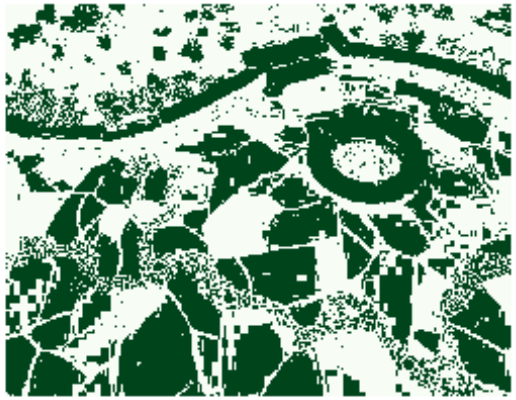
Outcome: An end-to-end processing pipeline.

Making use of: **17 million images** from **Google Street View** for 112 cities in the UK.

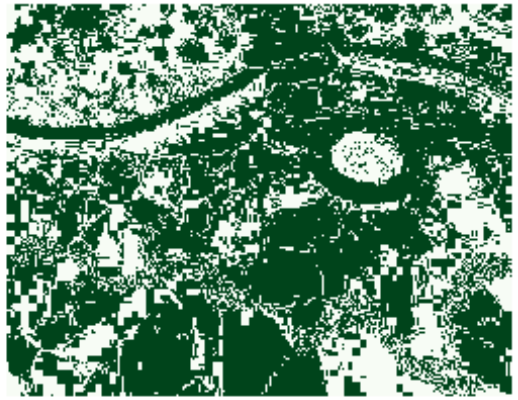
… **OpenStreetMap** road network data
… Deep **image segmentation** methods



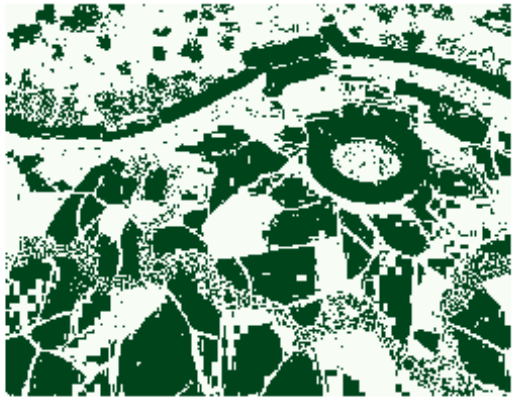
Aerial photography for green space statistics



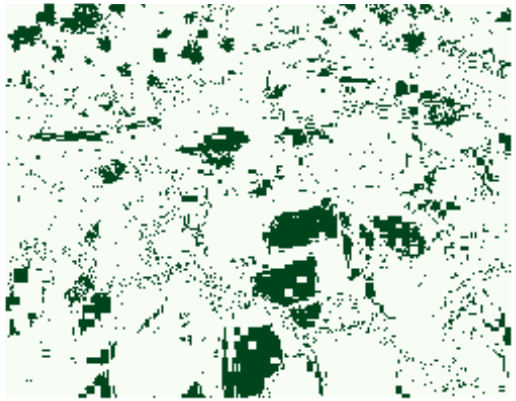
vNDVI (48.3%)



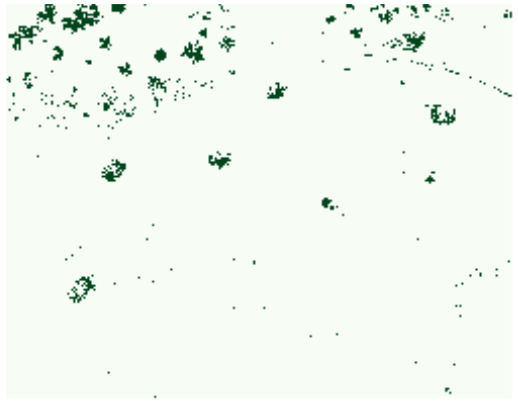
GLI (63.7%)



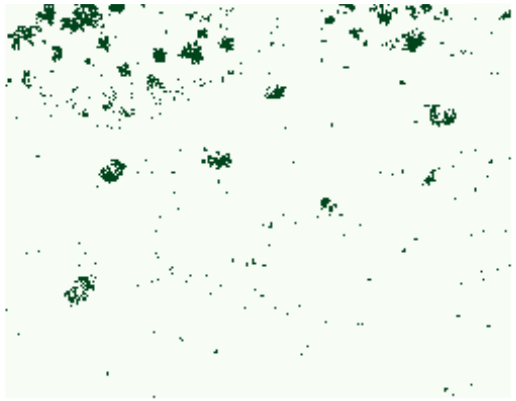
VARI (48.1%)



HSV (16.2%)



Lab(a*) (2.5%)



Lab(a*b*) (3.3%)

Data Science Campus

Growing public sector data science capability

Data4Good conference
16th October 2020

Tom Smith, Director
ONS Data Science Campus
@_datasmith