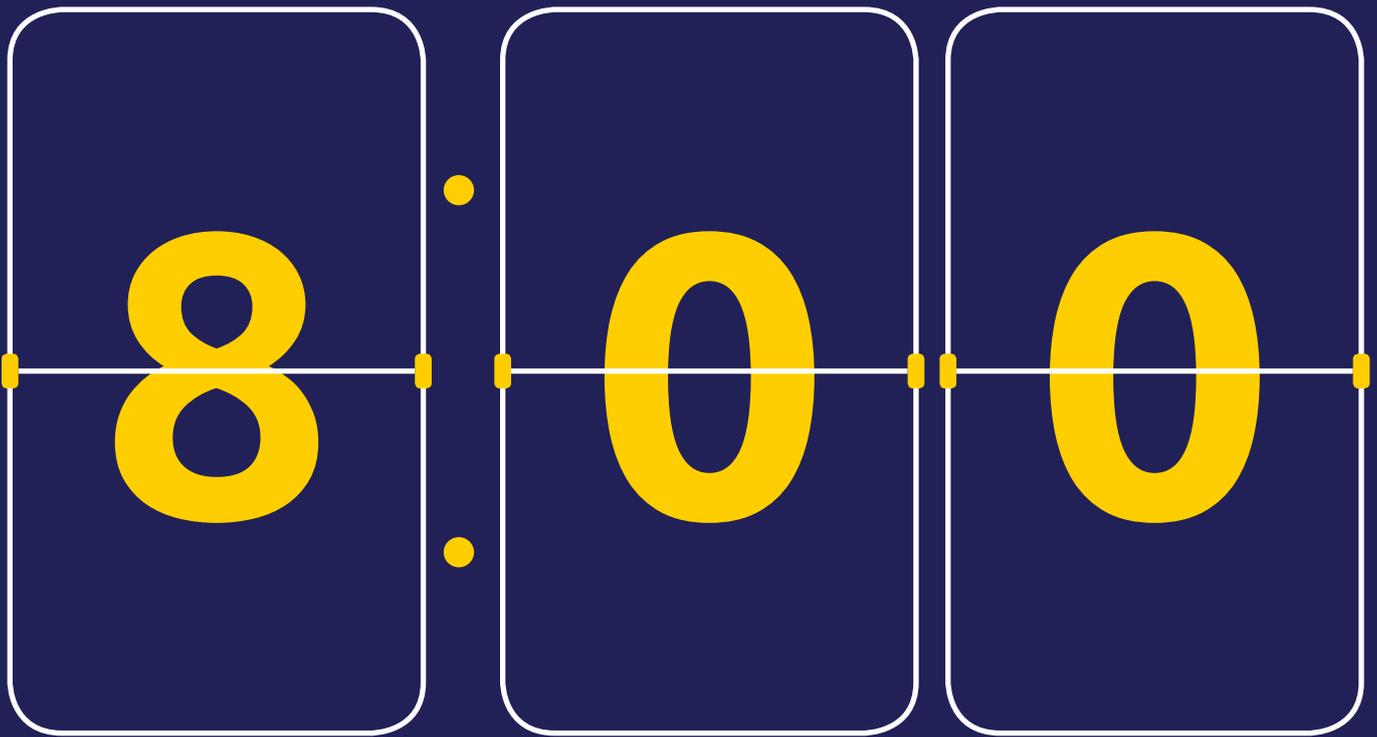


Data Bites

Getting things done with data in government



About this report

The Institute for Government believes that better data matters. Good-quality data is vital for both effectiveness – government needs to understand itself to improve its performance – and accountability – those of us on the outside are better able to scrutinise it.

Our work on data, primarily through our [Whitehall Monitor](#), [Performance Tracker](#) and [Parliamentary Monitor](#) projects (on central government, public services and parliament respectively), has thrown up a lot of things that could be better in government data. We collected some of these in a short report, *Gaps in Government Data*, in 2018.

But we also realised some very good work was going on, across various different fields related to data in government, and wanted to highlight these examples and bring people together around them. As a result, we launched Data Bites, our monthly meet-up showcasing interesting government data projects, in April 2019. This report summarises the first year of, and first eight, events.

 **#twitter:**
#IFGDataBites

Find out more:
[www.instituteforgovernment.org.uk/
data-bites](http://www.instituteforgovernment.org.uk/data-bites)

April 2020

Contents

About Data Bites	4
Introduction	6
Data Bites #1: April 2019	9
Data Bites #2: May 2019	17
Data Bites #3: June 2019	24
Data Bites #4: July 2019	29
Data Bites #5: September 2019	35
Data Bites #6: October 2019	41
Data Bites #7: November 2019	45
Data Bites #8: February 2020	53
Praise for Data Bites	57

About Data Bites

Better use of data is key to more effective government. Across government and the public sector, teams of public servants are doing fascinating work with data – on policy and public services, data sharing and visualisation, infrastructure and ethics, and much more besides. But those projects often don't get the attention they deserve.

'Data' means many different things across government – everything from statistics to service information to institutional memory – and is fragmented across (and within) different organisations, professions and functions.

For those not working directly with it, 'data' can still sound intimidating, prompting visions of cold, hard numbers in spreadsheet cells. But it need not be. Understood properly, the better use of data can bring huge benefits and help everyone across government make better policy, deliver better services and generally make things easier.

The Institute for Government's Data Bites events aim to breathe life into those numbers and make clear the benefits of using data in government.

A monthly meet-up consisting of four quickfire presentations, Data Bites aims to:

- showcase interesting government data projects
- bring together those in and around government with an interest in data
- demonstrate how better data can lead to more effective government.



Gavin Freeguard, IfG Programme Director, Head of Data and Transparency, and host of Data Bites.

Each of our presenters has just eight minutes, against the clock, followed by eight minutes of questions. (This is no coincidence: the basic unit of information is a byte, which consists of eight bits – eight is an important number for us!)

This report summarises the first eight (naturally) Data Bites events, held between April 2019 and February 2020. It draws out some key themes, and presents answers to a standard set of questions put to all of our speakers, allowing you to learn more about the 32 projects we've heard from to date. You can watch and listen back to all of the presentations on the [Data Bites](#) page of our website.

We are incredibly grateful to all of those sponsors who have supported the series so far:



If you would like to present at a future Data Bites event, or know someone who should, please contact gavin.freeguard@instituteforgovernment.org.uk

If you would be interested in supporting a future Data Bites event, please contact pritesh.mistry@instituteforgovernment.org.uk

All job titles/organisations correct as at time of presentation. Where speakers have been unable to provide answers, we have summarised their presentations.

Introduction

In the 12 months since our first Data Bites event, 36 speakers have given 32 presentations, each lasting eight minutes, across eight events, covering an incredible variety of subjects.

Our presenters have included civil servants in central government and other public bodies, public servants in local government, business people, charities, police officers, statisticians, data scientists, data engineers, data architects, auditors and many more. They have covered topics as diverse as artificial intelligence and data architecture, statistics and public services, ethics and infrastructure. Nonetheless, common themes have emerged. Here are eight.

1 Data sharing

The most prominent theme, touched on in every Data Bites event, is the need for more and better data sharing. Data sharing in government allows government to better tailor and target services and better understand the country it governs, the services it runs and the effect it is having.

This isn't just about data held by different departments and different tiers of government; Data Bites presenters from public and private sectors have wanted to throw their data at one another, given the insights into individuals and everyday life that could be generated by linking them. Although powers such as those granted under the Digital Economy Act 2017, guidelines including the government's Data Ethics Framework, initiatives demonstrating the value of linked data such as Administrative Data Research UK and new ways of thinking about data sharing like the Personal Data Exchange exist, there is still limited understanding of how to share data effectively, usefully and ethically.

This issue will only become more important: the Institute for Government and other civil society organisations have [previously called for](#) government to have the debate and discussion about how it uses citizens' data in public, with the public.

2 Data accessibility

A first step in being able to better share data would be to understand what data currently exists and be able to find it easily. But data is still too difficult to find. The Institute for Government has [recommended](#) departments publish and maintain a list of the datasets they are responsible for to combat this.

3 Fixing the plumbing

Data is much easier to use and share if it is of good quality. This includes different datasets consistently talking about the same things (whether organisations like departments, schools and local authorities, or individuals) in the same way (by adhering to standards) and using unique open identifiers to allow linkages; data being published in easier-to-use formats; and data not being locked up in old, legacy IT systems. But our presentations have shown these all remain significant problems.

There is a real data dilemma here, too: we can often only interest people in some of the difficult data infrastructure issues if we show what better data can mean in practice for their own lives and jobs, and often with exciting new technology; but that can make it much more difficult to focus on the difficult and often mundane 'plumbing' issues that need fixing to allow all of that to happen.

4 New types of data

The data world we live in has been profoundly transformed in recent years by increases in computing power (and storage), new technology and techniques, and new ways of collecting and using data. The very first Data Bites presentation – on how VAT returns, road traffic and ship movements could indicate how well the economy was performing more quickly than waiting for GDP statistics – very much set the tone for the opportunities available to government and society from new data sources.

5 User needs

The recent growth of digital government has brought more of a focus on service design and user needs. This also applies to data: how will different users and audiences want (and need) to use government data? This isn't just about formats and file types, but also how government communicates data and information. 'Users' should also include civil servants whose job brings them into contact with data – how can systems be set up to make it easier for them to collect and use data?

6 Seeing data as an asset

Data is an asset, but is more complex than more conventional assets, like property, that government will measure (as a [recent report](#) from the Open Data Institute and the Bennett Institute underlines). How can government properly regard data as a strategic asset, especially when so much of it is siloed between different departments and organisations, and understand the investment needed to improve its quality and use?

7 Dog-fooding

Type 'dog-fooding' into GOV.UK and you'll get 13 results: 11 actually refer to pet food. The [other two](#) are speeches from Matt Hancock in 2015 (when he was the minister responsible for data), which define dogfooding as 'using your own product' – apparently named after a pet food CEO who ate a tin of his product in front of shareholders. In data terms, it means departments publishing better open data because they are using it in their own day-to-day operations.

Though the word has barely featured through Data Bites, most of the presentations have focused on government organisations actually using data, and the importance of designing data into services and systems from the very start. If government organisations want others to use their data, and want to use data produced by others, the principle (if not the precise phrasing) remains vital.

8 Leadership and learning

The data agenda is a complex one, given it refers to so many different things in so many different fields, across so many different areas of responsibility. Government has, over decades, published numerous reports recommending improvements to problems which still exist many years on. It has not yet appointed the chief data officer promised by the 2017 Government Transformation Strategy, and Brexit and political uncertainty have delayed the National Data Strategy. It would be easy to detect (and become depressed by) a 'data malaise', as Adam Locker (Data Bites #6) puts it in this report.

Looking ahead

There are reasons to be more positive. As Data Bites has shown, there are some brilliant projects going on in government. There is a real desire from people to break out of siloes and share and learn from those projects. And there is real passion from those working on government, on data and digital and on particular service and policy challenges, for the better use of data. We hope the next eight Data Bites events will showcase that passion as well as the first eight have.

Indeed, government's use of data will only become more important in light of the [Covid-19 coronavirus pandemic](#) and government's handling of it. These events, and this report, were compiled before the current crisis, and therefore do not reflect recent developments. It is clear that the use of data will be a big theme in the response and aftermath of the outbreak, and this is something we hope to turn to in future.

Data Bites #1

April 2019

Sponsored by
VOCALINK
mastercard

Presentation	1:1
Topic	Faster Indicators of Economic Activity
Speaker	Louisa Nolan, Lead Data Scientist
Organisation	Office for National Statistics Data Science Campus

Can you summarise your project in a few sentences?

We have used big data sources and administrative data to produce faster, more timely indicators of UK economic activity.

What problem are you trying to solve?

The 2016 [Independent Review of UK Economic Statistics](#) stated that “the longer a decision maker has to wait for the statistics, the less useful are they likely to be”. With the growing availability of big data and large administrative datasets, and the tools, technology and skills to understand and process these, national statistics institutes everywhere are being challenged to produce outputs that meet the growing demand for more timely data.

The [Faster Indicators of UK Economic Activity project](#), led by the Data Science Campus at the Office for National Statistics (ONS), is an innovative response to the challenge of producing faster economic information.

What difference will it make to citizens?

Policy makers and analysts demand faster insight into the state of the UK economy in order to make informed, timely decisions on matters, such as the setting of interest rates, which affect the whole UK. The impact on the UK economy translates into the economic wellbeing of citizens, living standards, jobs, spending on services, etc.

How will you know when it has succeeded?

It's possible we may never know! If faster indicators (and likely other indicators as well) inform monetary or economic policy



to avert an economic downturn, it may be difficult to identify. However, we accept success as seeing these indicators included in the suite of information and insight used to inform not just policy decisions, but also the decisions of businesses and investors.

What are the main hurdles?

Using data collected for a different purpose by a third party carries the risk of changes to data collection, methodology, publication etc., which you are not in control of. It is really important to collaborate well with the data collectors and relevant subject matter experts to understand these, and to properly understand any quality issues.

If you could change one thing about data in government, what would it be?

The Campus would have successfully completed [our synthetic data programme](#), and made it possible to produce good-quality synthetic microdata (record-level data), which preserves privacy requirements, for a wide range of useful datasets.

Since Data Bites...

Since our presentation at the inaugural Data Bites session in April 2019, we have been publishing a regular monthly update of our three sets of indicators, based on VAT returns, road-traffic counts for England and ship-tracking data. We have also created a more timely output of shipping indicators, published weekly, that includes indicators for cargo ships and tankers, and which removes non-trade-related shipping such as pleasure craft and ferries.

We have had much interest in these outputs. Policy makers and economists have been monitoring the UK economy, and internationally, we have been working with the UN AIS (Automatic Identification System) Task Team and UN Global Platform to build international capability in the use of the AIS ship-tracking data.

Key links

[Faster indicators of UK economic activity](#) (article)

[Faster indicators of UK economic activity: more timely and relevant shipping indicators](#) (article)

[Synthetic data for public good and art](#) (blogpost)

[Research Output: Economic activity, faster indicators, UK](#) (monthly report)

[Weekly shipping indicators](#) (weekly report)



Presentation	1:2
Topic	Data in housing and land
Speaker	Paul Maltby, Chief Digital Officer
Organisation	Ministry of Housing, Communities and Local Government (MHCLG)

Can you summarise your project in a few sentences?

Digital Land is a small multidisciplinary team in MHCLG. We're working to make more land and housing data available for the emerging property technology ('PropTech') sector.

What problem are you trying to solve?

The [housing marketplace](#) is failing. As with many market failures, that's in part due to a lack of information; not enough people know where they can build a house, or how to go about building. The few people who can navigate the planning system or have access to information about land have more power in negotiations with the communities and others over development. As we've seen with other sectors, digital services have the potential to revolutionise the housing market. There doesn't seem to be a shortage of investment for new companies who want to build services to inform decisions and remove inefficiencies throughout the housing market.

But this innovation is stifled by a lack of data. Where data does exist, it isn't good enough for people to build services upon. It's currently quite hard to discover what data is needed for any given decision. Many of these decisions are devolved to local planning authorities, which can be the local authority, a national park, a development corporation, or in the case of neighbourhood planning, the parish. Finding data from some of these organisations is hard, but bringing them together to provide a service nationally is much harder. When you do find a dataset, it's often in a different format to other publishers, and requires work to build it into a service. It's then hard to navigate often restrictive and complex licensing terms, and it's hard to trust the data will be kept up to date. Finally it's hard to know how to report any issues you may spot with the data, and have them resolved at source.

What difference will it make to citizens?

Better data will help people find more information about housing where they're already looking: on property websites, mapping services and online communities. The nascent PropTech sector will be able to provide services to better inform planning authorities when negotiating with developers, and help them to better engage with communities to decide where development should take place. Improving data will help developers make better decisions when targeting investment. Making planning policies, rules and site information available as data will enable drawing packages that better inform architects of the constraints when designing buildings, and it will be much simpler to buy a modular building 'off the shelf', and be certain it can be placed on your plot.

How will you know when it has succeeded?

The scale of the problem we're tackling is enormous; for example, a planning application may need to reference more than 350 datasets as evidence. Nationally this data is held by around 400 different public bodies, and often the same data is commissioned and collected multiple times. Our approach so far has been to work on individual projects to help policy makers have better outcomes from data standards. Working on these individual projects is how we are learning how to work at scale, with many policy teams and organisations at once, to build a national picture of development.

We're taking an evolutionary approach, meaning the value of our work emerges as we go along. However, we believe the value of our work grows more quickly because of the links between the data we make available: links means there's a network effect. There may be a point when we could claim to have had a revolutionary impact, but meanwhile we'll have a series of successes along the way.

What are the main hurdles?

Making data available is seen by many authorities as an externality; we are asking busy people to take time and effort to make data available for others. We hope that making it easier to publish data as a part of their existing processes will reduce costs, and managing feedback on the data will mitigate many of the perceived risks. We also believe that making services present their data in a national context will help them to see the value of publishing their data.

If you could change one thing about data in government, what would it be?

Government is currently quite well served with statistics, but most services need more precision. They need unique identifiers for public statements such as a site proposed for development, the company making a planning application, the addresses of the houses being built and their Land Registry titles. These identifiers don't always exist, are hard to find, or are restricted in their use. We should help the people responsible for making decisions throughout government to record them consistently in registers with persistent, unique, open identifiers. This is the heart of the vision for [GOV.UK registers](#).

Key links

[MHCLG digital](#) (blog)

[Digital land](#) (blogposts)

[It can be difficult to make data from information](#) (MHCLG blogpost)

*Answers provided by Paul Downey,
Head of Digital Land, MHCLG*



Presentation	1:3
Topic	Using data to uncover new, increasingly sophisticated patterns of financial crime
Speaker	David Divitt, Vice President, Financial Crime Product Management
Organisation	Vocalink

Can you summarise your project in a few sentences?

We've been working with financial institutions, payments schemes and industry groups to examine the behaviour of 'money mules' as they steal and move money across a real-time payments system. We wanted to learn how they do it, in terms of the characteristics of the movement of money and the accounts involved. And we wanted to leverage artificial intelligence (AI) to build a solution that would detect and prevent them.

What problem are you trying to solve?

We are trying to help solve the problem of financial crime (fraud and money laundering) in account to account payments systems.

What difference will it make to citizens?

The more we can do to stop fraudsters and subsequent money launderers, the less they will be able to steal from people, businesses and governments – an issue that is estimated to cost up to \$2 trillion per year. The harder we make it for them to launder their proceeds of crime, the less attractive it will be for them to steal money in the first place.

How will you know when it has succeeded?

The ultimate measure of success is seeing fraud and financial crime in certain areas fall over time as new techniques have been put in place to stop it – and we are seeing this happen. Measuring success directly to this one single solution can be a challenge, as it is an additional source of information for participating banks which works alongside their existing anti-fraud and anti-money-laundering solutions.

There is no single silver bullet to solve this issue.



What are the main hurdles?

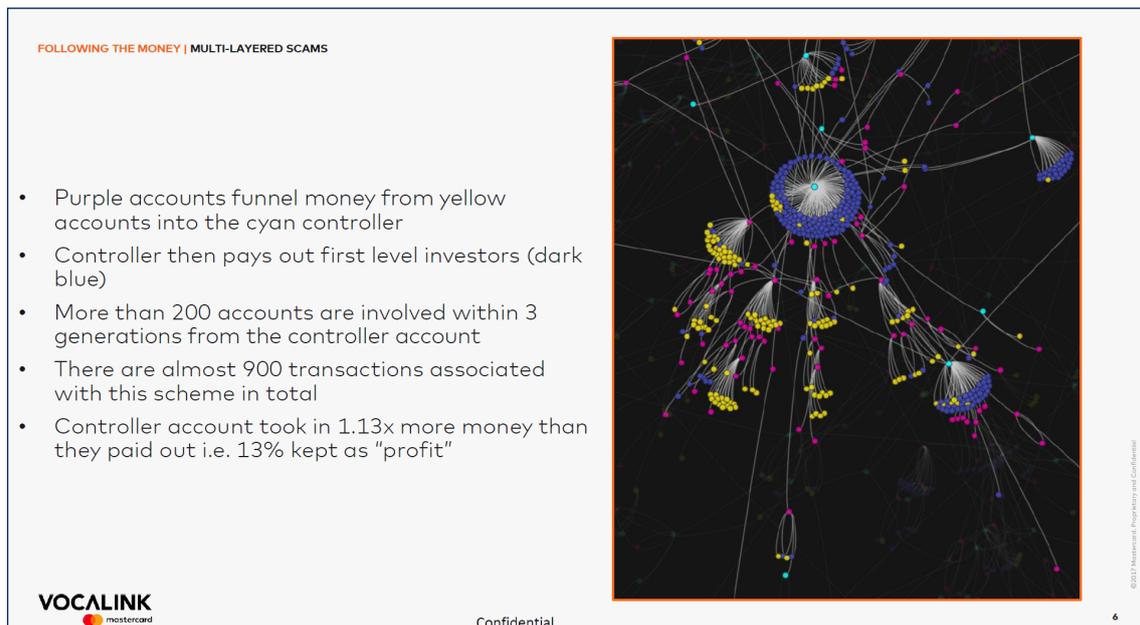
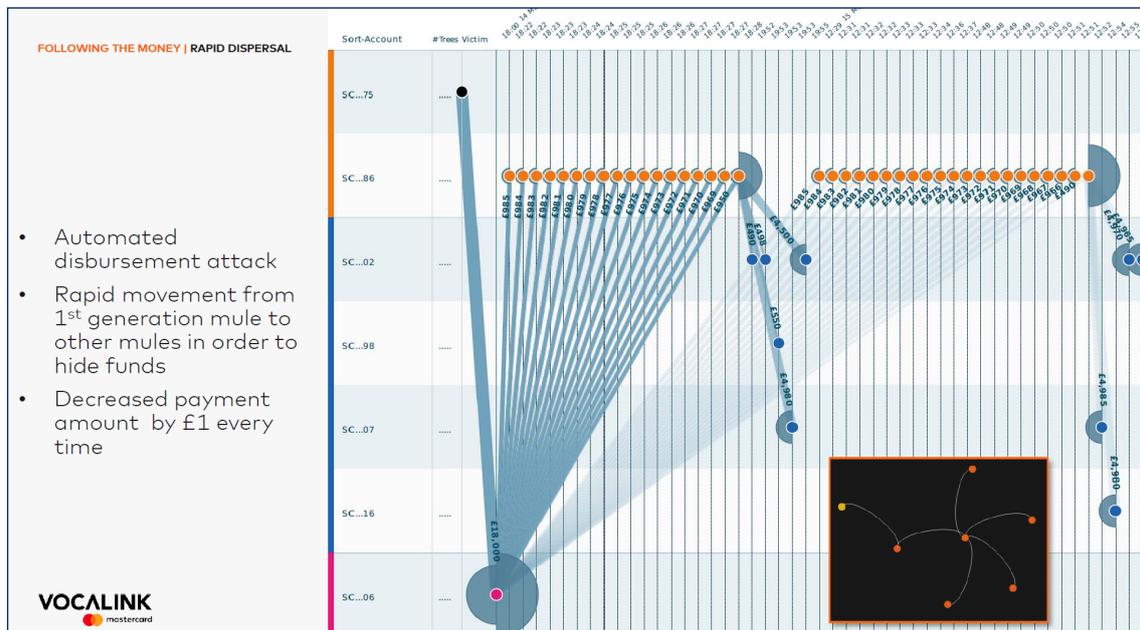
Criminals change their approaches all the time, so adapting quickly to those changes is key. In the time between running the pilot and Trace Financial Crime going live, around 18 months later, we saw a dramatic difference in financial crime across the UK payments network. New patterns of fraud had emerged, and the speed at which stolen funds were being split and dispersed across mule accounts [increased]. Continuing to invest in the capabilities that drive this solution is a key area of focus for the team.

If you could change one thing about data in government, what would it be?

Quite often it is a problem of not knowing what is out there and available, and where to get it. If there was a way to have a central catalogue of data it would make knowing what is available much simpler.

Key link

[The Rise of the Mule](#) (report)



Presentation	1:4
Topic	Data analytics at Ofgem
Speakers	Sophie Adams, Product Owner Jon Downing, Data Scientist
Organisation	Office of Gas and Electricity Markets (Ofgem)

Can you summarise your project in a few sentences?

Ofgem has the opportunity to improve data transparency and create new insight for the transforming UK energy market, while reducing the data burden on industry. By using modern data-management tools and governance approaches, Ofgem can change and grow greater trust in regulatory data and better-informed exchanges with stakeholders through deeper insight.

By making data and methodologies transparent and accessible, Ofgem can drive accountability in industry, collaborate in new ways with its stakeholders and innovate.

What problem are you trying to solve?

There are many problems surrounding data including the lack of transparency, auditability and governance of data in regulated sectors. We are approaching this in a holistic way, learning how we can work better with industry and become more open and transparent.

What difference will it make to citizens?

We are working to create principles around using data across industry including the use of data within government. We are enabling better data sharing and accessibility that in turn will lead to better data quality and sharing of this information. Industry will be able to better help consumers who may need a little extra support.



How will you know when it has succeeded?

This is an ongoing process, and we are working incrementally, however one particular success is when we can share data across sectors to answer some of our big questions.

What are the main hurdles?

We are working on gaining buy-in across the sector. While we have support, there is also funding required. Without both of these we cannot succeed. There are other hurdles in each area of data services; however we are overcoming these as they come up.

If you could change one thing about data in government, what would it be?

Accessibility! There is a huge amount of data across government, but it is often hidden and not fully utilised.

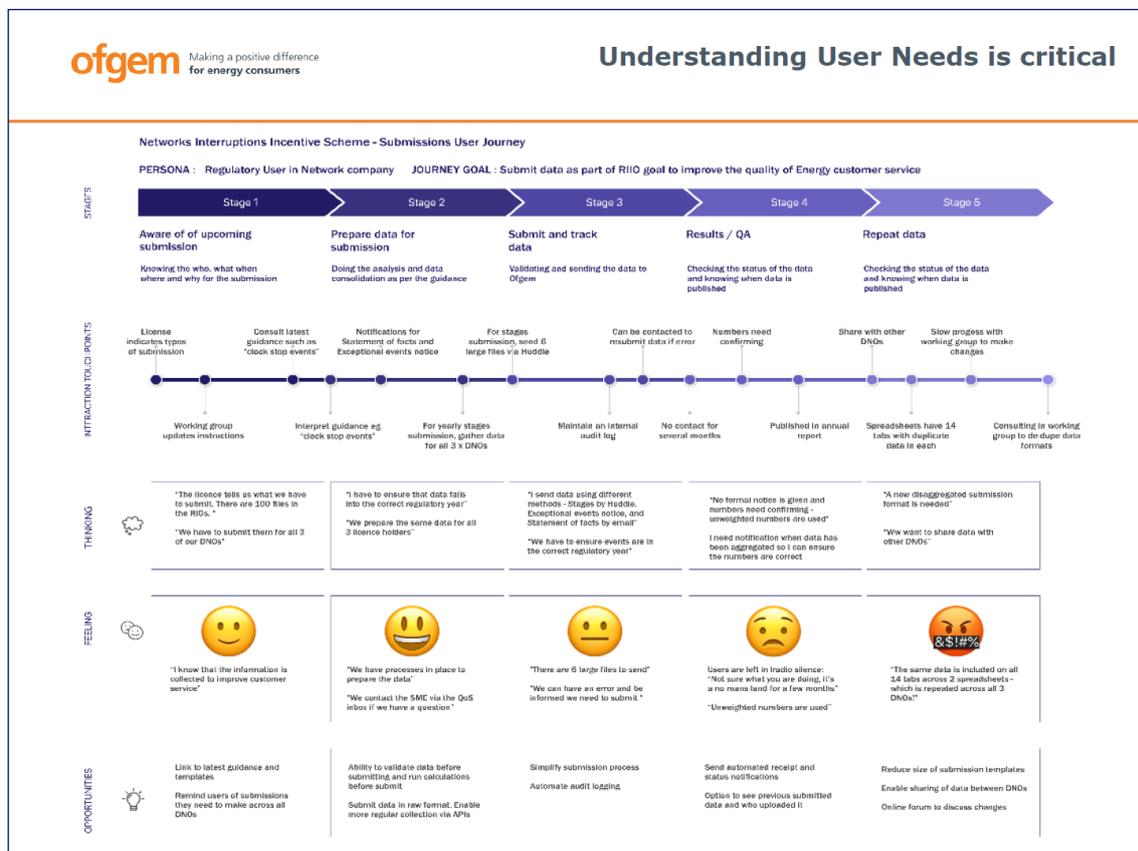
Since Data Bites...

The platform has continued to develop. The [Data Exchange](#) is live and streaming data from industry into Ofgem's data platform. We are working to [open source our code](#) for others in government and industry to use. We continue to support transformation of the energy sector through the [Modernising Energy Data](#) workstream.

Key links

[Ofgem code on GitLab](#) (platform)

[Modernising Energy Data](#) (workstream)



Data Bites #2

May 2019



Sponsored by

Presentation **2:1**

Topic **Bringing skills demand data to local government via a digital service**

Speaker **Aleks Bobrowska, Data Scientist**

Organisation **Department for Work and Pensions Data Science Newcastle Hub**

“Securely obtaining and working with data from different government departments will be the biggest hurdle ahead.”

Can you summarise your project in a few sentences?

We created a prototype of a digital information service for local government, to provide insights about the digital skills market in their area. This involved a lot of advanced analytics and working with big data, as well as gathering user feedback and designing an interactive web application.

What problem are you trying to solve?

Local government organisations, such as LEPs [Local Enterprise Partnerships] and local authorities, don't have consistent access to data about the skills and jobs in demand in their local area.

Without that clear picture it's hard for them to influence employers and training providers to target the right areas and people with the right training. This creates a skills gap between the jobseekers and the demand on the market.



What difference will it make to citizens?

As local government has access to appropriate and consistent evidence on the skills and jobs in demand, they will be able to influence training providers to deliver vital education to the citizens. With the right skills, people will be able to fill the local jobs that are in demand now as well as in the future. Having enough digital skills will also enable people to progress to better employment.

How will you know when it has succeeded?

When our users are happy and report using the tool to extract evidence from data to create and support their local industrial strategy.

What are the main hurdles?

Securely obtaining and working with data from different government departments will be the biggest hurdle ahead.

If you could change one thing about data in government, what would it be?

If we could share data between departments easily, we could build a much more rounded and complete picture of the society, the economy and the environment we live in, allowing us to better serve citizens.



Presentation **2:2**

Topic **The UK government's Data Ethics Framework**

Speakers **Sam Roberts, Head of Open Government and Open Data
Sarah Gates, Head of EU Exit and International Strategy
(Government Office for Science)**

Organisation **Department for Digital, Culture, Media and Sport**

The Data Ethics Framework

In 2016, the government data science partnership commissioned work on [public dialogues](#) around the use of data – the public were wary about some uses, but happier if these were transparent and well managed. Government started to think about guidelines that would translate the plethora of law and ethics into practical advice that could communicate risks and benefits easily to stakeholders, and produced the first framework.

Government then began to rethink the framework, talking to a range of users and moving away from a focus on data science to talk more holistically about data, including information assurance, security and practicalities. The seven principles of the framework are to:

- start with a clear user need and public benefit
- be aware of relevant legislation and codes of practice
- use data that is proportionate to the user need
- understand the limitations of the data
- ensure robust practices and work within your skillset
- make your work transparent and be accountable
- embed data use responsibly.

It is aimed at “anyone working directly or indirectly with data in government”. There is a workbook published alongside it which people and teams can use for self-assessment. It has already been adopted by others (including the Department of Health and Social Care, and government guidance on artificial intelligence). Work continues to understand how everyone is using it, to develop more practical guidance and to understand how it is changing how data projects are being managed.

Key link

[Data Ethics Framework](#) (guidance)



Presentation **2:3**

Topic **Personal Data Exchange**

Speaker **Simon Worthington, Founder, Register Dynamics (formerly of the Government Digital Service)**

Organisation **Register Dynamics**

“Making [data] more visible is a great place to start.”

Can you summarise your project in a few sentences?

Personal Data Exchange (PDE) is a data-sharing federation that makes personal data sharing for government services easier, more private and more secure. It aims to make what one department knows available to another department's service, sharing only what needs to be known for eligibility and making sure citizens' data isn't used for any other purpose. It provides audited, compliant, privacy-preserving and useful data access everywhere, both online and for staff.

What problem are you trying to solve?

Government is currently very far from 'once-only' – where citizens tell government something once and have that information available everywhere. Instead, citizens must collect information from around government, mainly in the form of bits of paper, and scan or send it manually. This is a lot of work and puts the burden of proving eligibility on the citizen. Citizens often get it wrong, and correcting for these mistakes is a huge cost for services. For services, having a live link to fix this is currently very expensive – PDE aims to make it cheaper and safer.

What difference will it make to citizens?

It'll make applying for a government service easier. Instead of having to scan originals or fill out forms, citizens can only submit basic identity information and have government work out their eligibility automatically – with confidence that this is done in a transparent and privacy-preserving way.



How will you know when it has succeeded?

When lots of services start using it – and when services start to collect less information!

What are the main hurdles?

Departments that hold data are very risk-averse: they don't give out their data lightly, and don't want to lose control. They are also concerned that they will be asked to provide information for a purpose that it shouldn't be used for. Thankfully, PDE keeps control of both of these things with them – but the challenge is convincing them of this.

If you could change one thing about data in government, what would it be?

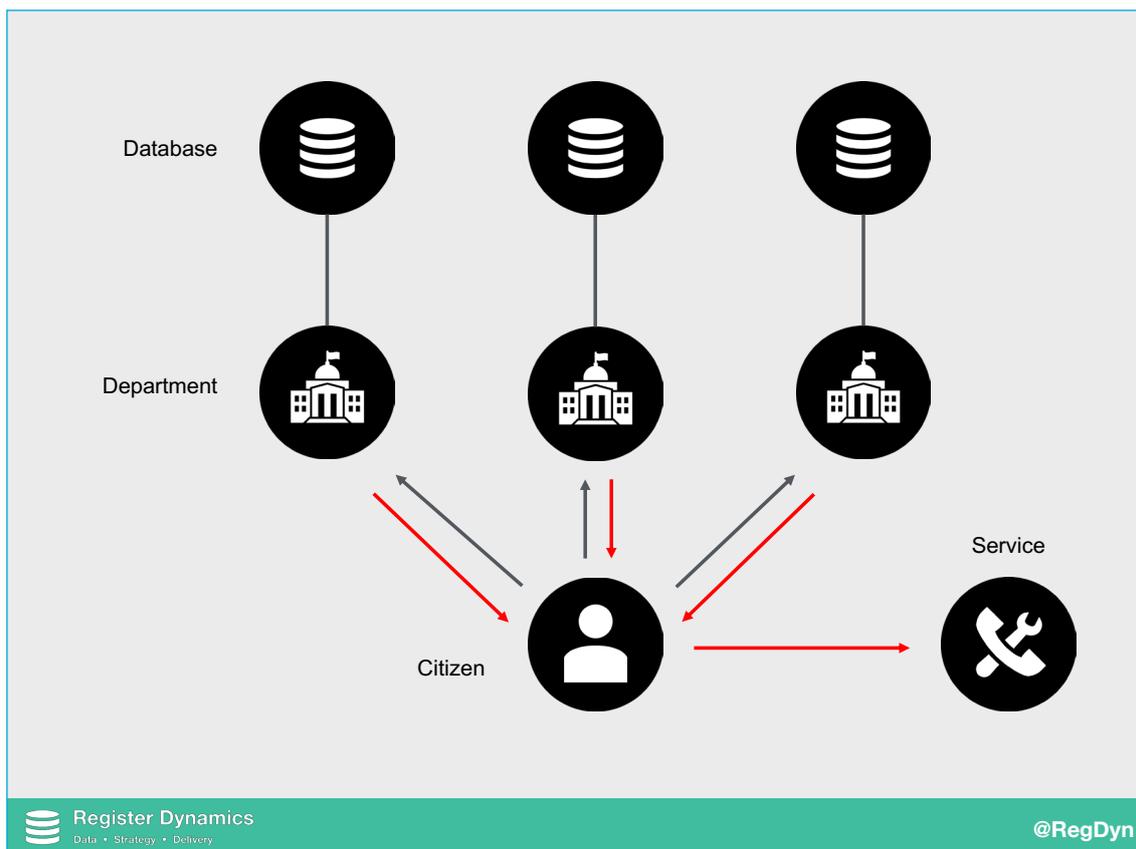
Wow, only one thing?! I honestly think making it more visible is a great place to start. If we had a complete picture of what data government collects and who is responsible for it, that'd go a long way to fixing a lot of problems – even just working out who to contact to start setting up a data share can take small departments many weeks of toiling!

Key links

[Digital eligibility checks for service teams](#) (blogpost)

[Attributes and Personal Data Exchange](#) (product teams) (all public information from GDS)

[Aqua: Personal data for cross-government services](#) (article)



Presentation	2:4
Topic	The Noise App
Speaker	Martin Pilkington, Director Huw Williams, Business Development
Organisation	RHE Global

“There is caution about data sharing in government.”

Can you summarise your project in a few sentences?

RHE Global is a leading provider of apps to local government, and as The Noise App and Reportable become widely established we aim to work with the relevant government bodies to analyse the data gathered to inform policy decisions. This will assist in reducing environmental pollution such as noise, anti-social behaviour and environmental crime, generating positive social outcomes.

What problem are you trying to solve?

We ensure data is gathered speedily by the citizen and provided to government scientists and experts. We standardise data formats and are developing advanced analytics which will be done in strict compliance with GDPR [General Data Protection Regulation] and information governance regulations.

What difference will it make to citizens?

Citizen reporting to the government via apps has already improved the process for both these parties in relation to reporting noise nuisance. With predictive analytics some problems can be prevented before they happen, improving the health and wellbeing of citizens.

How will you know when it has succeeded?

We will see a reduction of the need to report and resolve issues if we are successful. This will improve the environment, improve health and wellbeing and reduce costs for the government.



What are the main hurdles?

The main hurdles are cultural change, educating the citizen and the speed and spread of related smartphone technology.

If you could change one thing about data in government, what would it be?

There is caution about data sharing in government. If the correct procedures are used to safeguard privacy then increased co-operation by government agencies to resolve social and environmental problems could be facilitated by speedier and standardised data sharing.

Key links

[The Noise App](#) (blog)

[The Noise App](#) (website)

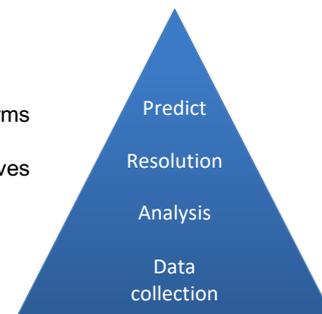
[Seven successful government apps from around the world](#) (article)

The Noise App

Simplify noise nuisance reporting and investigation

Overview

- This native App exemplifies **citizen science reporting**.
- Noise Nuisance is >50% of all anti-social behaviour and is the **second largest source of environmental pollution**
- Triage/process transformation tool – prioritise the real problems
- Standardise digital reporting replaces written diary sheets or web forms
- Reduces the stress of reporting noise nuisance for residents and saves officer time/cost
- Improves overall quality of evidence provided-can be used with standard noise investigation equipment
- Manages cases efficiently through www.thenoiseapp.com website
- Enables **cross functional cooperation** social services/police/environmental health



© 2019 RH Environmental Ltd. trading as RHE Global

www.rheglobal.com

Data Bites #3

June 2019

Presentation	3:1
Topic	The work of the Geospatial Commission
Speaker	William Priest, Chief Executive
Organisation	Geospatial Commission

The Geospatial Commission

Geospatial data is data about place and location: “everything is somewhere”. Eighty per cent of all datasets are location or place-based. The UK has a ‘plethora and history’ of the world’s best geospatial data, and is rated second in the world by the Global Geospatial Readiness Index.

The Geospatial Commission is an expert committee at the heart of government. A 2017 manifesto commitment, it was set up in autumn 2017 with £80m “to do good things [to] leverage and realise both economic value and public sector value” of geospatial data and “optimise public services”. It provides strategic oversight for six partner bodies: the Valuation Office Agency, Ordnance Survey, HM Land Registry, the Coal Authority, the UK Hydrographic Office and the British Geological Survey.

It estimates that £6–11bn per annum in incremental value could be unlocked if geospatial data could be brought together and made more accessible to government departments, the wider public sector and the private sector. The Commission aims to drive quality as well as accessibility, oversees some of the government’s key geospatial contracts, and has already embarked on initiatives including:

- a £1.5m crowdsourcing competition for innovative projects
- £3.9m for two pilot projects mapping underground assets
- consolidating several licences into a single Data Exploration Licence
- making parts of Ordnance Survey’s MasterMap freely available.



Presentation	3:2
Topic	Data and the web
Speaker	Andy Dudfield, Chief Publishing Officer
Organisation	Office for National Statistics (ONS)

Can you summarise your project in a few sentences?

Making statistics available online in a consistent manner, with consistent metadata and consistent labels.

What problem are you trying to solve?

Too much statistical data runs on inconsistently formatted [Microsoft] Excel and this makes working with official statistics harder than it should be. The projects discussed were looking to address this for both humans and machines consuming information and allow for easier joining of data from multiple sources.

What difference will it make to citizens?

It should open up statistics to a wider audience than ever before.

How will you know when it has succeeded?

Fewer datasets are downloaded from the ONS site (because people don't have to guess what is in them) and over time, fewer government websites publishing statistics.

What are the main hurdles?

Everything runs on Excel and this is a complex domain requiring deep domain expertise across a lot of specialist areas.

If you could change one thing about data in government, what would it be?

I would publish all of the statistics in one place.

Key links

[Why we have produced a Beta version of a new ONS API \(blogpost\)](#)
[ONS Digital \(blog\)](#)



Presentation	3:3
Topic	How data science has allowed BEIS to tackle old problems in new ways
Speakers	David Fry, Chief Statistician Luke Perera, Head of Data Science
Organisation	Department for Business, Energy and Industrial Strategy (BEIS)

“[BEIS] had looked at lots of economic problems over the years, but had never had enough data before.”

Data science at BEIS

Two years ago, BEIS’s small data science team was doing a lot of good work, much of it below the radar. They decided they needed a strategy to make the most of the expertise they were building. This involved developing and maintaining skills, acting as a consultancy for everyone across the department, building a data science community and ensuring they had the right technical infrastructure.

The main trick, though, was picking a few projects that allowed them to engage with senior people in the department.

The department had looked at lots of economic problems over the years, but had never had enough data before. Now there is a huge ‘data exhaust’, which can allow things to be tracked indirectly – for example, we may not be able to track the diffusion of new technology, but can look at what jobs are being advertised and how these are changing. There is now much more data, different types of data, faster data processing, and new tools and techniques.

Recent projects have used machine learning – including one with the ONS Data Science Campus analysing patents – and natural language processing, such as looking at job adverts in different regions and how they’re changing over time. They also have an award-winning data science project mentoring scheme, which has helped them understand the art of the possible – it is still early days for ‘data science’ – and build a community.



Presentation	3:4
Topic	The challenges in using data in government
Speaker	Yvonne Gallagher, Director of Value for Money, Digital
Organisation	National Audit Office

“Government should recognise and address the need to tackle the underlying data ‘plumbing’ issues first.”

Can you summarise your project in a few sentences?

Many believe that government can do much more with its data, for example: more open data; data sharing; personalisation of services; analytics and cutting-edge technology; reduction in the citizen burden. These are all good aims. They sound easy to achieve – but they are not as there are significant blockers and constraints which will require sustained effort to overcome.

What problem are you trying to solve?

Getting senior leaders to understand why it is so important to address data shortcomings and the need to prioritise and fund the work.

What difference will it make to citizens?

More efficient services; more trust in government data; more accurate holding of customer data; reduction in the citizen burden to provide and prove data over and over.

How will you know when it has succeeded?

This is an elephant which will need to be bitten off in chunks. It needs a sustained commitment over a significant period of time so that:

- the quality of the data in the underlying systems has been improved
- detailed processes and procedures reflecting the aims of each function have been put in place around privacy, security and consent at working level for staff so they know what is/isn't allowed
- inefficient processes to correct and support the use of data of inadequate quality are no longer normalised and are seen as costed problems to be addressed
- agreed cross-government key data standards.

What are the main hurdles?

There are several:

- Government does not treat data as a strategic asset.
- Funding pressures can inhibit progress on data projects.
- Data quality is often inadequate.
- A lack of data standards across government has led to inconsistent ways of recording the same information.
- Legacy systems often only work for the policy they were built to deliver so data inconsistency is built in.
- Historical ways of working can inhibit progress.

If you could change one thing about data in government, what would it be?

Government should recognise and address the need to tackle the underlying data 'plumbing' issues first.



Data Bites #4

July 2019

Sponsored by



Office for
Statistics Regulation

Presentation	4:1
Topic	Statistics in the age of data abundance
Speaker	Catherine Bromley, Statistics Regulator
Organisation	Office for Statistics Regulation

Can you summarise your project in a few sentences?

We are the UK's independent regulator of official statistics produced by government and the NHS. We expect statistics published by public sector bodies to be produced in a trustworthy way, be of high quality, and to provide value by informing answers to society's important questions.

What problem are you trying to solve?

We now live in a world of increasingly abundant data and statistics producers need to adapt to this environment by innovating and maximising the potential use of data for both citizens and organisations, and by making data available for wider reuse with appropriate safeguards.

What difference will it make to citizens?

Greater access to trustworthy information helps citizens to make informed choices about their lives, and to understand the world in which they live.

How will you know when it has succeeded?

When we stop needing to highlight areas for producers to improve. When we see data being joined up to create better statistics efficiently and effectively.

What are the main hurdles?

Poor user engagement that fails to recognise user needs, or over-prioritises the needs of specific users at the expense of others. Risk aversion around data sharing. Poor leadership.



If you could change one thing about data in government, what would it be?

It would be seen as a common shared resource for the public good, rather than as the property of particular departments or agencies.

Since Data Bites...

We've published the final version of our vision statement (we distributed the draft at Data Bites) and published an update to the *Joining Up Data for Better Statistics* report. I have written new regulatory guidance for official stats producers who provide data for researchers (working title 'Unlocking the value of data through onward sharing'). It sets out our expectations of producers about data standards and quality, data provision to users, and developments to official statistics (as a result of insights from secondary analyses). It will be shared for comments with some data providers and users who we know have an interest in this issue before publication in the spring.

Key links

[Joining Up Data for Better Statistics](#) (report)

[Joining Up Data for Better Statistics](#) (blogpost)

[Joining Up Data for Better Statistics – 2019 Update](#) (progress report)

[How to stop the bad data driving out the good](#) (blogpost)

[Statistics that Serve the Public Good](#) (vision)

[Building Confidence in the Handling and Use of Data](#) (guidance)

[Code of Practice for Statistics](#) (guidance)

Office for Statistics Regulation

UK Statistics Authority

Office for Statistics Regulation

Systemic Review Programme

Joining Up Data for Better Statistics

September 2018

Data sharing and linkage help to answer society's important questions

Analysts have the skills and resources needed to carry out high-quality data linkage and analysis

Government demonstrates its trustworthiness to share and link data through robust data safeguarding and clear public communication

Data sharing decisions are ethical, timely, proportionate and transparent

Data are documented adequately, quality assessed and continuously improved

Project proposal assessments are robust, efficient and transparent

Presentation **4:2**

Topic **Newham's Data Warehouse**

Speaker **Dr Jane Kennedy, Head of Research and Data Hub**

Organisation **London Borough of Newham**

The Data Warehouse

Before introducing the Data Warehouse, Newham had 30 different systems that didn't integrate, didn't speak to each other, and it was very hard to make operational decisions using the data they held. The Data Warehouse creates a single view of people, events, property, of children within a school and the local business sector. This isn't just effective from an operational perspective and for making savings, but for data science, machine learning and predictive analytics.

Benefits include immediate improvements to data quality – fundamental to be able to link data across services – identifying errors and detecting fraud. The council can make better operational decisions; for example, the housing service can now look at other data about a tenant and understand why they may not be paying rent (for example, due to financial problems or mental health issues) and support the tenant in a way they couldn't before.

They've also been able to improve strategic decision making, restructured services (for instance, around looked-after children), made savings, looked at cost avoidance and income generation, and generally helped services to meet their statutory obligations more effectively and efficiently.



Presentation **4:3**

Topic **Using machine learning to help users find content**

Speaker **Ganesh Senthil, Senior Product Manager, GOV.UK**

Organisation **Government Digital Service (GDS)**

“We are exploring whether we can use data science to better understand user journeys, behaviour and their intent on GOV.UK.”

Can you summarise your project in a few sentences?

We wanted to investigate whether we could build a recommendation engine that provides users with relevant and contextual onward journeys as related links on content pages.

What problem are you trying to solve?

Around 98% of the content on GOV.UK does not have any related links which means a user can find themselves in a dead-end. The only way of moving onwards is either through the header, footer or search.

What difference will it make to citizens?

For citizens, we can now provide relevant and contextual onward journeys for users on all content on GOV.UK.

How will you know when it has succeeded?

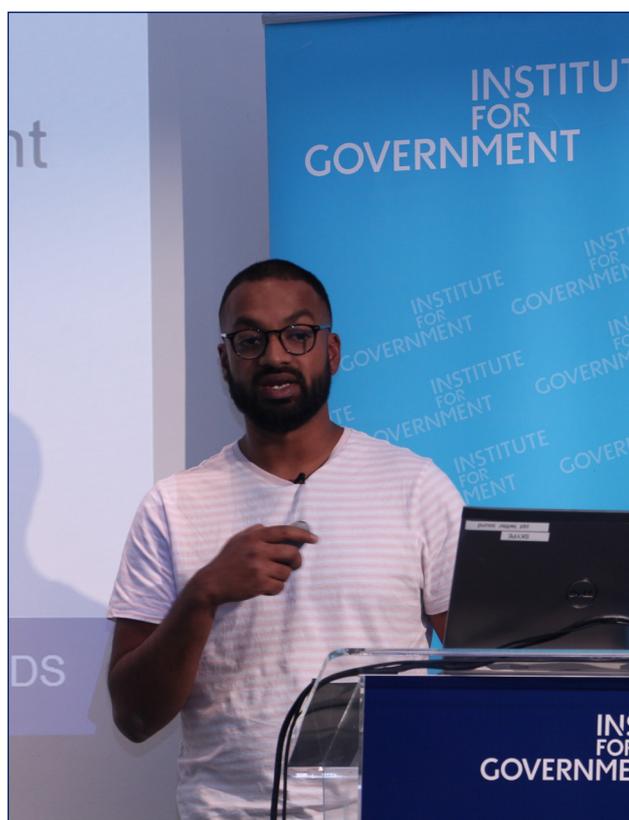
We will see a decrease in users clicking on the header, footer and using search, and an increase in satisfaction on our user survey.

What are the main hurdles?

Ensuring links are accurate and relevant to the content that a user is on.

If you could change one thing about data in government, what would it be?

Data silos. In order for a user to complete a task, they often have to visit multiple different organisations, but we as digital teams do not share the data between us to gain a full picture of what a user is doing.



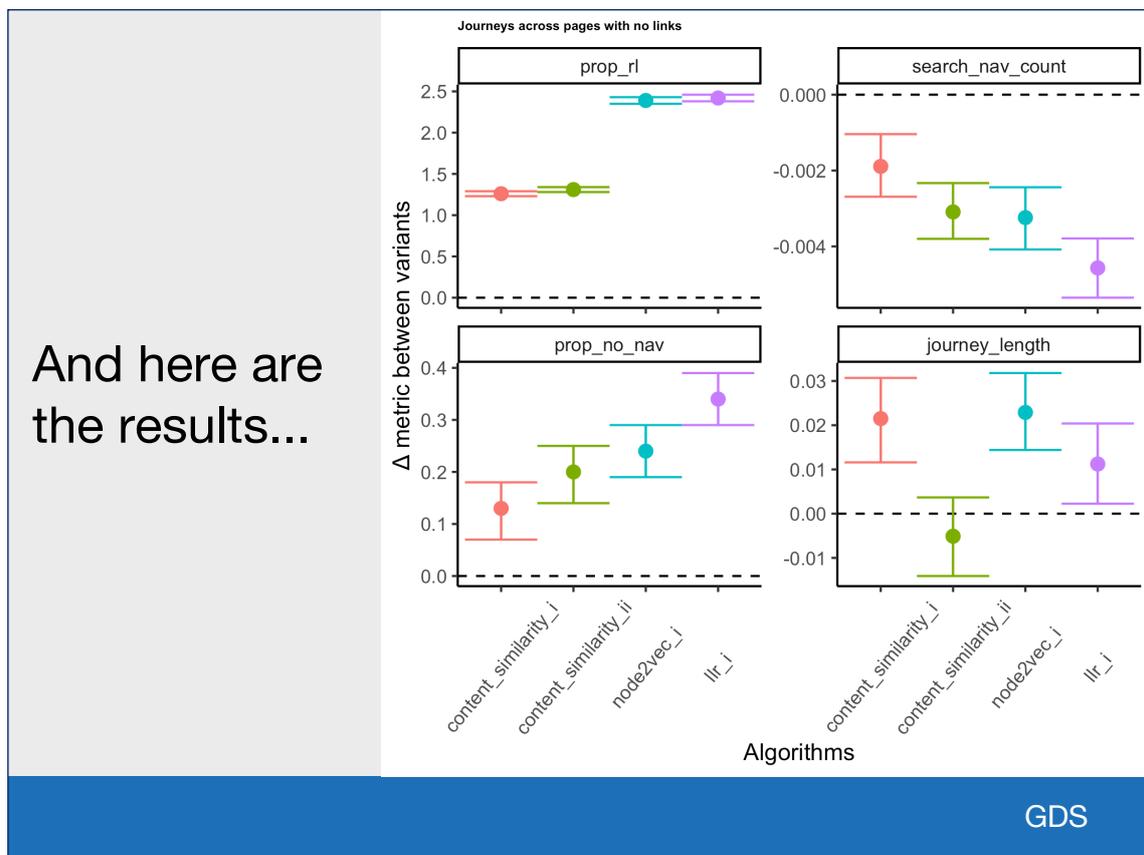
Since Data Bites...

Shortly after I presented at the event, I formed GOV.UK Data Labs and I think all the things we are working on would be of interest to Data Bites (I hope!). For example, we are using Natural Language Processing to structure and enrich our content, we are using graph databases as an analytical tool and we are exploring whether we can use data science to better understand user journeys, behaviour and their intent on GOV.UK. There is a blog post on our work going out shortly, which will give a better idea of the what and why.

Key links

[Training algorithms to create related content links \(blogpost\)](#)

[When worlds collide: putting data science into production \(blogpost\)](#)



Presentation	4:4
Topic	The AI Barometer
Speaker	Michael Birtwistle, Policy Adviser
Organisation	Centre for Data Ethics and Innovation (CDEI)

The Centre for Data Ethics and Innovation

CDEI was set up to advise government on maximising the benefits of artificial intelligence (AI) and data for society, to understand how to develop and deploy technology ethically and safely, and understand the risks and opportunities of AI and how prepared governance systems are to mitigate or take advantage.

The centre is developing a number of outputs, including major reviews on [algorithmic bias](#), and on [online targeting and profiling](#), and short 'snapshot' papers on areas of topical public concern (including [deepfakes](#), [smart speakers and voice assistants](#), and [AI and insurance](#)). Other reports will explore options in more detail with a view to making recommendations to government, such as a forthcoming report on [data-sharing models](#).

The AI Barometer project

The AI Barometer looks across the whole landscape and is designed to help policy makers understand what needs attention. It seeks to help understand the risk, opportunities and governance gaps around every application of AI and data-driven tech, including which are the most important and urgent, in every context and sector.



The range of risks and opportunities is obviously very broad, and trying to map out the impact of algorithmic decision making on individuals, on physical harms, on confidence in institutions, on market effects and much else besides is obviously a challenge. There are many different lenses to look at all this through. CDEI are hoping to build a typology or hierarchy of these risks and opportunities, and help build a common picture and language for talking about them.

Data Bites #5

September 2019

Sponsored by

EQUIFAX[®]

Presentation **5:1**

Topic **Using data to identify vulnerable service users and treat them more fairly**

Speaker **Oli Vogel, Head of Recovery Performance**

Organisation **Indesser**

Can you summarise your project in a few sentences?

We are trying to use data and analytics to improve outcomes for vulnerable customers who are in debt.

What problem are you trying to solve?

Currently, vulnerable customers are identified only when we speak to them; but what if we could identify them sooner, and alter our treatment of these individuals accordingly?

What difference will it make to citizens?

The difference it will make to citizens is a more informed collections strategy for vulnerable customers.

What are the main hurdles?

The hurdles, from a data perspective, are that definitions of vulnerability are somewhat subjective, and that not all vulnerable customers are known to us. Research shows many do not engage with the debt process.

If you could change one thing about data in government, what would it be?

If I could change one thing it would be that data sharing, for the benefit of citizens, is built into how government operates.



Presentation	5:2
Topic	Breaking down the barriers to administrative data research
Speaker	Emma Gordon, Director of Strategic Hub
Organisation	Administrative Data Research UK (ADR UK)

“A greater shared understanding that administrative data is an invaluable resource for public good.”

Can you summarise your project in a few sentences?

ADR UK’s work plays an important role in bridging the gap between government and academia, enabling government policy to be informed by the best evidence available, and putting us on the path to a future in which the true potential of administrative data to improve society is realised.

What problem are you trying to solve?

Good government policy – policy that solves social and economic problems and improves lives – needs good evidence. But good-quality evidence is expensive and time-consuming to produce, and evidence from academic research can also have a different focus from what government needs most to make policy decisions and deliver effective services. ADR UK’s mission is to solve both these problems by transforming the way researchers access the UK’s wealth of public sector data.

What difference will it make to citizens?

ADR UK is driven by a mission to stimulate and enable research that not only enhances our understanding of how UK society works but also has the power to bring about meaningful change. Through the research we enable, we are able to gain a better understanding of UK society, with the ultimate aim of informing policy decisions that have the potential to improve lives.

How will you know when it has succeeded?

The ADR UK programme is currently funded across the entire partnership until March 2021. We will know if we have succeeded if the work we are funding in this investment period builds into a body



of knowledge great enough to demonstrate a shift in thinking about how government works with external researchers, in time for a further investment decision to be made that allows the momentum we have built to continue.

What are the main hurdles?

For government departments to see opening up access to administrative data for research via secure settings as standard practice, rather than the exception.

If you could change one thing about data in government, what would it be?

A greater shared understanding that administrative data is an invaluable resource for public good.

Since Data Bites...

We published our [annual report](#) on the same day as Data Bites, summarising the first year of the investment. The growing strength of administrative data research internationally was highlighted by the [4th International Conference on Administrative Data Research](#) we sponsored in December. We announced a grant of nearly £3m to the [Ministry of Justice](#) for their data-linking programme, 'Data First', in March 2020.

Key links

[ADR UK 2018–19: Reflections on our first year](#) (annual report)

[Harnessing the potential of linked administrative data for the justice system](#) (blogpost)

[A vision for ADR UK](#) (article – forthcoming)



Linking Census and educational outcomes data for England

ADR UK is currently working with ONS and DfE to link data from the 2011 Census with a bespoke extract of the National Pupil Database (NPD).

  UK Research and Innovation

Presentation	5:3
Topic	Transforming capability in data and analytics to deal with public sector fraud
Speaker	Graeme Thomson, Programme Director, Counter Fraud (Data Analytics, Capability Development, and Pilots)
Organisation	Cabinet Office

“Fraud is big. Fraud is really big. I mean it’s £48bn worth of big. It’s a department, it’s MoD [Ministry of Defence]. And it’s so hard.”

Data and counter-fraud work

Fraud is complex, and there is no single panacea for tackling it across government. But why weren’t people in government using data to combat fraud? It was because some departments didn’t recognise they had fraud and had no analysts; some recognised they had fraud, but had no analysts; and some recognised they had fraud and had the analysts, but still couldn’t quite bring things together.

There is little understanding of data as an asset in government. There are legislative challenges, so they changed the law – the Digital Economy Act includes a power to share data to counter fraud. This got rid of arguments about “the Data Protection Act stops me” (although it never did) by granting an explicit power. Government also can’t take a private sector solution and apply it to government; unlike the private sector, it can’t turn ‘clients’ away.

Graeme’s team have focused on specific fraud problems and run small, agile pilots to tackle them. They have brought together people interested in tackling fraud, and tried to pass their learning on. They are in the game of fishing – teaching others how to do it for themselves – and have developed a best practice guide, workflows and links with the capability matrix (at which stage should you involve analysts, which ones, etc). Challenges remain – fraud is expensive, you can’t simply prosecute your way out of it, it isn’t high up the agenda compared to other types of crime. But they are focused on fraud prevention, developing capability across government through linking into the cross-government profession, developing toolkits and guides (around data mindset, capability, ethics, access and quality) and have [published a thought paper](#) for external comments.



Presentation **5:4**

Topic **The future of data at the Ministry of Justice (MoJ)**

Speaker **Sam Tazzyman, Head of Data Engineering and Analytical Platform**

Organisation **MoJ**

Can you summarise your project in a few sentences?

Government in general (and MoJ specifically) aims to use the large amount of data generated by our services to improve how we do things. To do this we need to improve the way data flows through our organisation. Data engineers in the MoJ use cutting-edge technology to radically alter the way data is retrieved from source databases, cleaned, reshaped and served to analysts and data scientists, and then displayed to operational staff and senior decision makers. This opens up our data to investigation and means we can find out ways to improve things for everyone!

What problem are you trying to solve?

Our legacy data flows have usually developed in an ad-hoc way, and are complicated and full of manual unaudited steps. In particular, this means that the datasets on which we make decisions can be out of date, difficult to audit, inconsistent between copies and very hard to reproduce. Analysts and data scientists have to spend time and effort overcoming data issues, rather than having the data available to them.

What difference will it make to citizens?

Any part of the criminal justice system can potentially be improved by insights from our data. Some potential examples include improvements to safety in prisons, better court scheduling and better answers to parliamentary questions. Policy can be developed on the basis of clear evidence because data will be much more available to test hypotheses. Additionally, a lot of analysts' time is currently spent sourcing, cleaning and preparing data. Our work will automate this, so that our analysts and data scientists can spend more time on their actual work, and produce more insights.



How will you know when it has succeeded?

We will have specific examples in prisons and courts where we will be able to point to our data making a difference for operational staff (we already have some of these). MoJ's statistical releases, and internal forecasts, will be largely automated. Data supporting policy decisions will be auditable and reproducible so that decisions can be revisited and justified where necessary.

What are the main hurdles?

The databases are large, complicated and difficult to access because often the storage and maintenance of them is outsourced. The cultural change of moving people to new ways of working can be difficult.

If you could change one thing about data in government, what would it be?

Improved understanding about how valuable data can be if it's accurate and accessible. This applies at all levels, from operational staff filling in the data in the first place, to senior leaders making decisions about outsourcing data storage and access. The biggest companies in the world see data as their biggest asset: if the public sector does likewise, we could make great improvements to how things are done.

Key links

[Bringing disparate data sources under control with good metadata](#) (blogpost)

[Pushing the boundaries of data science with the MOJ Analytical Platform](#) (blogpost)

[Engineering the data of the future](#) (blogpost)



Data Bites #6

October 2019

The
Legal
Education
Foundation

Sponsored by

Presentation	6:1
Topic	Legislation as data – meeting the challenge of Brexit
Speaker	John Sheridan, Digital Director
Organisation	The National Archives

The National Archives, data and Brexit

Legislation is data. Law is both legislation, and case/common law. A piece of legislation changes law from one state to a new state. What glues data together on legislation.gov.uk is URLs (web addresses): the URL scheme is “a symphony to the power of URLs for allowing you to bring together document-oriented information and data”, such as changes over time. This comes into its own when facing a challenge like the European Union (Withdrawal) Act 2018, which is like turning the power off for a whole body of law. It deals with the consequences of repealing the European Communities Act of 1972 by incorporating law that directly applies to the UK into the domestic statute book.

The National Archives aimed to support government departments with legislating for exit, fulfil their duties under the Withdrawal Act, and aid legal certainty for the public. It rolled out two services – a web archive for EU exit legislation, and incorporating EU legislation onto legislation.gov.uk. Domesticating EU law is a big transformation job – you have to really understand the data in order to transform it from the EU data model to a UK model.

Bringing together EU documents, UK documents, EU data about amendments, and UK data about amendments allows you to perform interesting analysis of the impact of EU exit on our statute books.

In order to prepare for multiple scenarios before 31 October 2019, legislation.gov.uk built a ‘data clutch’ between the two datasets: in the case of a no-deal exit, they could disengage the clutch and apply UK amendments; in the case of a deal, they could flexibly shift gear and synchronise where they needed to (goods in Northern Ireland) and disengage where they didn’t.



Presentation	6:2
Topic	Digital Justice: HMCTS data strategy and delivering access to justice
Speaker	Dr Natalie Byrom, Director of Research and Learning
Organisation	The Legal Education Foundation

HMCTS and Digital Justice

Natalie launched her new report which included a 29-point plan for putting access to justice at the heart of the data strategy being developed by HM Courts and Tribunals Service (HMCTS) as part of its digital transformation (a £1bn reform programme is of unprecedented scale, scope and speed). It's vital that people can understand and use the law, so decisions about reforming processes must be based on good data and robust evidence. The judiciary has set out principles, including that reform should maintain or improve access to justice.

In 2018, Natalie was seconded as expert adviser to HMCTS on open data and academic engagement. Her key findings, on what data needed to be collected to ensure the new digital processes comply with the Equalities Act 2010 and ensure access to justice, included the need for better data on individuals in the system, better data on whether processes deliver access to justice as legally defined, and (most importantly) better data on outcomes (including who drops out of the system and why).

This really matters: for the vulnerable, at the sharp end of government decision making, the justice system is a safety net. If you bring an appeal against the government's refusal to award you a personal independence payment, there is a 73% likelihood that you will succeed – meaning 73% of original decisions were wrong.

With the exception of legislation.gov.uk, the wider data landscape around the justice system has been described as 'dysfunctional', with opaque arrangements for accessing data, a lack of data asset registers, lack of a common language for talking about justice data, and a lacuna of ownership and leadership. There is an urgent need for access to justice impact assessments for reformed services, to develop (and compel) good practice in collecting demographic and equality data, and to raise the profile of the issues raised by the report.

Key link

[Digital Justice: HMCTS data strategy and delivering access to justice \(report\)](#)



Presentation **6:3**

Topic **Fix the plumbing: a user's guide**

Speaker **Adam Locker, Data Architect**

Organisation **Food Standards Agency**

Can you summarise your project in a few sentences?

Less of a project and more of a cathartic expression of lessons learned from several years trying to 'fix the plumbing' in a small government agency, and an attempt to put some cheer back into the data community by reminding them how far we've come.

What problem are you trying to solve?

All of them; bad data, slow data, inconsistent data, data of dubious origin. Making sure we're talking about the same thing across our business.

What difference will it make to citizens?

At the risk of going overboard, if the practical solutions we're finding to problems might help other government departments and agencies make the same gains then the difference could be profound and also unseen. Might stop some folk moaning.

How will you know when it has succeeded?

When the (proven false) statistic about 80% of a data scientist's time being spent cleaning data is only 60%, and also proven to be false.

What are the main hurdles?

Motivation, co-ordination, participation. I worry that there's a data malaise setting in again with the data agenda being a bit all over the place. If I've learned anything about data in government is that the people who care about it, *really* care about it, and are not to be trifled with.

If you could change one thing about data in government, what would it be?

A little less telling, and a little more showing. Events like Data Bites are key to that.

Key link

[Data as institutional memory](#) (blogpost)



Presentation **6:4**

Topic **The National Audit Office's (NAO) Data Service**

Speaker **Ben Coleman, Data Analytics Lead**

Organisation **NAO**

The NAO Data Service

The 'bright and shiny' NAO Data Service deals with some of the issues with how government releases data, allowing it to more easily use it in its work (financial audit, and undertaking value-for-money inquiries). The fact it needs a local 'solution' implies there are problems: 'this bit [of the presentation] could go on forever'.

These problems include:

- The inconsistent use of identifiers: the names of prisons are different across different Ministry of Justice datasets; the names of local authorities are different across departments; the names of government organisations are different for different years in the same single statistical series. This makes it more difficult to link data.
- Difficult formats: spreadsheets with page breaks and merged cells, optimised for printing not analysis. Time-series spread across many releases and files, in separate parts of GOV.UK. Data as text.
- Revisions not always being well publicised.

All of this means more time is spent cleaning the data (the oft-quoted statistic says 80%) rather than analysing and using it.

In 2015, the NAO started building its solution: the Data Service, an internal data warehouse. It contains 500 statistical measures, drawn from published datasets and overseen by a central team, who do the cleaning and conforming centrally ensuring common dimensions (prisons, local authorities, etc.), with scripts and an evolved set of synonyms. It is an interactive web-based app (built in R), supported by a sharepoint site with metadata. The project started small, agile and has been built over time. The NAO is looking to add new features, such as helping people across the organisation add things in themselves.



Data Bites #7

November 2019

Sponsored by



Presentation	7:1
Topic	Geospatial's role in better citizen outcomes
Speaker	Miranda Sharp, Innovation Director
Organisation	Ordnance Survey

Can you summarise your project in a few sentences?

Geographic data has a multiplicity of uses, many in government but also in the industries and sectors that deliver service and public good. It is a powerful tool to engage users and decision makers in information through attractive presentation on a map, enabling them to derive new insight and drive evidenced-based, measurable action. Examples range from the creation of the beautiful river network, the lurid land-use change statistics and analysis of the impact of the Tadcaster bridge collapse. (Or, alternatively, analysis of property prices following infrastructure investment.)

What problem are you trying to solve?

Nothing less than the better organisation of information, by using place, so that the best context-specific and relevant information is available to people making decisions about physical assets.

What difference will it make to citizens?

The best information in the right hands provides a platform for human flourishing. Benefits are not just lower consumer cost, less waste and more certainty when planning new transport investments but also the ability to tackle cross-silo challenges such as achieving net zero carbon, tackling poor air quality and increasing productivity.

How will you know when it has succeeded?

We already have seen many examples in the use of geographic data making a difference for people, prioritising



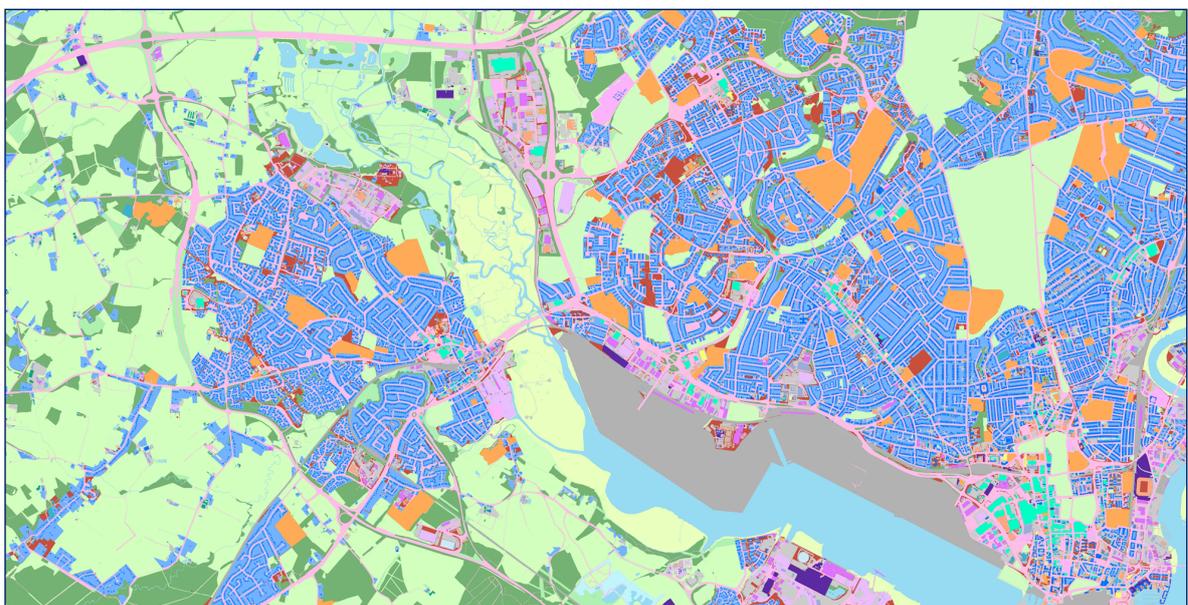
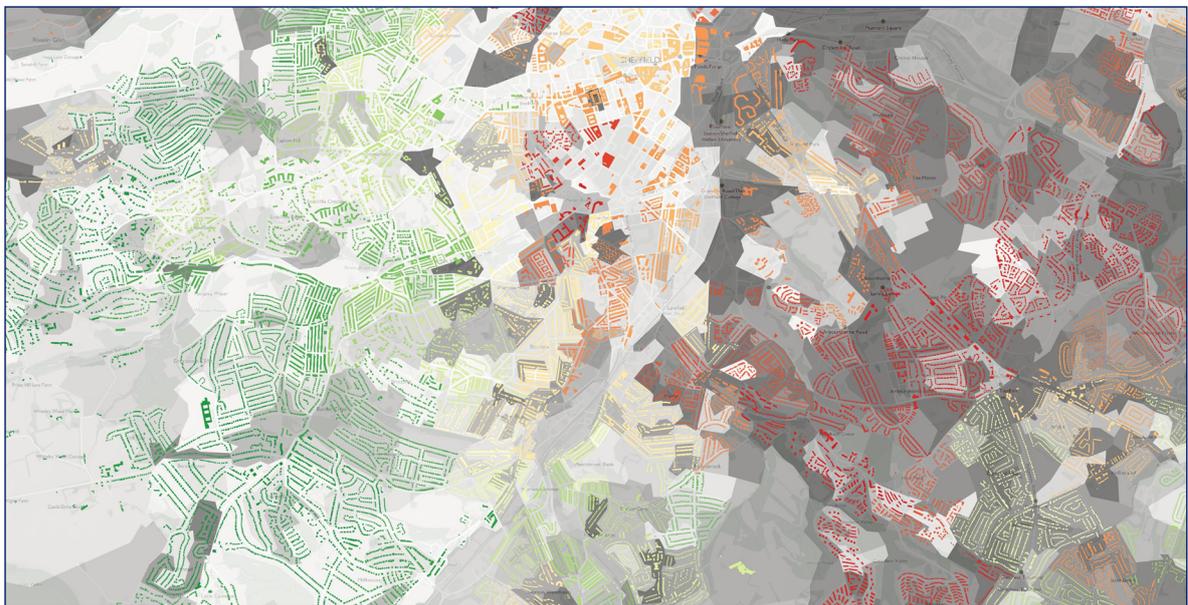
coastguard resources and saving time in emergency response. However, there is a great deal more to do and we hope that by working with the Geospatial Commission and others that more projects and departments will benefit from the use of high-quality geographic data.

What are the main hurdles?

Historically, geographic data has been difficult to use and access. We hope that recent licence changes (Discovery and Evaluation licence), OS Zoomstack and the new arrangements under the PSGA [Public Sector Geospatial Agreement] will enable more teams to engage.

If you could change one thing about data in government, what would it be?

For the National Data Strategy to be about all data, not just personal data and for the National Infrastructure Strategy to consider data or 'e-infrastructure' as well as the physical environment.



Presentation	7:2
Topic	Digital Excellence – using data as a catalyst for creating economic value
Speaker	Nic Granger, Director of Corporate
Organisation	Oil and Gas Authority (OGA)

Can you summarise your project in a few sentences?

The Information Management programme was designed to maximise the value derived from petroleum-related data. This has been achieved by the combination of the creation of new statutory powers on data stewardship and publication, the creation of a new and important piece of digital national infrastructure – the UK National Data Repository (NDR) – and the execution of work to digitalise the OGA, leading to a new digital capability, the OGA Digital Energy Platform.

What problem are you trying to solve?

Lack of easy access to large quantities of available petroleum-related data was hampering the oil and gas industry. By making this data available to all it is now being accessed in 191 countries worldwide and is leading to new and novel uses of the data as well as advanced technologies, such as artificial intelligence and machine learning.

What difference will it make to citizens?

The widespread availability of data for the offshore petroleum industry means that it can be used to optimise hydrocarbon production and exploration and thereby holds the potential to increase efficiency and economic value for the UK. The data is also being used in energy-integration projects as we transition to a low-carbon economy.

How will you know when it has succeeded?

The NDR commenced operating in March 2019 with strong support from industry, academia and the supply chain. It is early days but the usage of the data in the system has increased by 100% with very positive feedback across users. The data is now being combined with other data and analysis obtained by the OGA in the OGA Digital Energy Platform to provide authoritative insight and analysis and ensure effective stewardship of hydrocarbon production activities and the move to energy transition.



What are the main hurdles?

The main hurdle has been changing culture towards using data in a collaborative and transparent way.

Key links

The Oil and Gas Authority launches one of the largest ever public data releases (announcement)

Industry dives into UK's new oil and gas data hub (article)

UK NDR six month anniversary (blogpost)

OGA Digital Platform Users

 Oil & Gas Authority

70,000
visitors to OGA's website with the express purpose accessing data



183TB
of data downloaded from OGA digital platform



Users from
184
countries have used the OGA digital platform



750k+
page views across the OGA Digital Platform



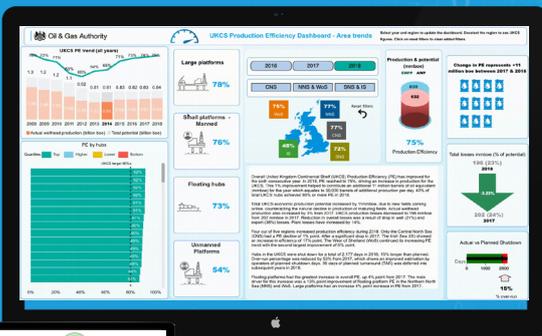
“ Science thrives on the open exchange of information and ideas, and I commend the Oil and Gas Authority for making a wealth of valuable data freely available.”

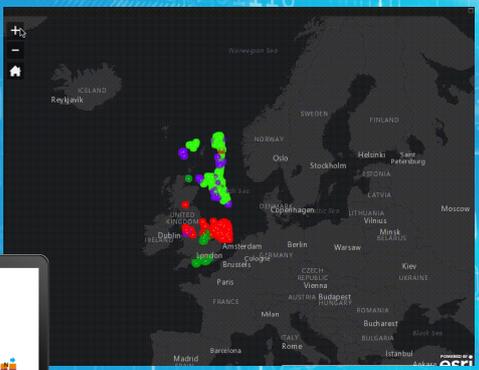
Professor George Boyne, Principal, University of Aberdeen

4578
registered users of NDR (60% public)

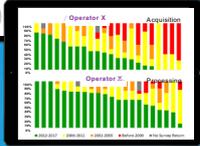


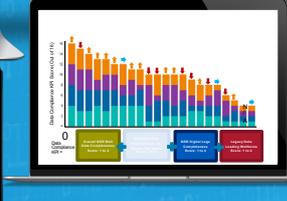
Freely available to all











Putting data at the centre of decision making and value creation



Presentation	7:3
Topic	The potential of charity data
Speaker	David Kane, Product Lead
Organisation	360Giving

Can you summarise your project in a few sentences?

360Giving helps funders share data about what and who they fund in an open, standardised format. Sharing data in the 360Giving Data Standard makes it easier to see and understand UK grant making. Funders, researchers and wider civil society can access, compare and use the data to learn more about the sector and make better funding decisions.

What problem are you trying to solve?

Funding often takes place in isolation, without the sort of sophisticated data that would be available in the commercial sector. Funders can feel like they are 'giving in the dark' without the context they need to make sure their money is spent wisely. Equally, grantees find it hard to find out if a funder is a good fit for them, to inform their fundraising. Researchers and policy makers also find it hard to understand the true size and scale of the sector and its changing needs.

What difference will it make to citizens?

We believe better data about grants leads to better funding decisions and more strategic grant making. Improving grant making means improving outcomes for beneficiaries – the people and organisations who ultimately benefit from the funding. Being able to see government grants side by side with grants from local authorities and independent trusts and foundations makes it easier to understand the different flows of funding to civil society, to identify changes over time and to hold funders accountable for their decisions.

How will you know when it has succeeded?

When sharing funding data in this way becomes the norm, and the data is widely used across the voluntary sector and beyond. But our goal is not just about the grants data covered by the 360Giving



Data Standard – we want grant making and the voluntary sector as a whole to become more data-informed and to have a culture where data and evidence are at the heart of decision making and learning.

What are the main hurdles?

Grant making is often not at the forefront of technological change, and some funders are reluctant to spend money on themselves to improve their data and systems, so some of our work is making the case for that investment. We're aware of the challenge of producing consistent data across all government departments, so we've provided guidance on how it can implement the Data Standard in its systems and what a good quality grants dataset should look like.

If you could change one thing about data in government, what would it be?

We want the publication of grants in a standard format to be a normal part of how government operates. It has approved the 360Giving Data Standard as a government standard and needs to embed this into its grant management systems. The second step is to use that data, along with data published by independent funders, to help guide better funding decisions and help understand the changing needs and demands of UK civil society.

Key links

[360Giving](#) (website)

[GrantNav](#) (website)

[360Insights](#) (website)

360 INSIGHTS

SEE YOUR GRANTMAKING
in new ways

Combine and visualise 360Giving, charity and other data to answer questions about your grantmaking

Select a 360Giving dataset
or upload a file that meets the 360Giving Standard.

CHOOSE A DATASET

insights.threesixtygiving.org

Upload a

Presentation	7:4
Topic	Using open and shared data to improve local public services
Speaker	Miranda Marcus, R&D Programme Lead
Organisation	Open Data Institute

“Better use of data can streamline processes, provide information to citizens, and make data available to innovators and service providers.”

Can you summarise your project in a few sentences?

Local governments provide essential citizen services but are often swamped by demand in an environment of increasing populations, budget cuts and limited resources. We’ve been researching how open and shared data can be used to improve local services, working with councils and service designers around the UK, and have developed a number of resources, tools and guides to help service providers with their work.

What problem are you trying to solve?

We’re trying to lower the barriers to using data in local public services by creating simple, non-technical tools which address the barriers that service delivery managers commonly face when trying to use data.

What difference will it make to citizens?

Better use of data can streamline processes, provide information to citizens and make data available to innovators and service providers. It can save money, improve existing services, enable new services, support better policy decisions and empower citizens to have more informed conversations with their local authority.

How will you know when it has succeeded?

In the long run, if there are more services successfully using open and shared data. In the short run, people using the toolkit and giving us feedback.



What are the main hurdles?

There are several, including a lack of:

- data skills and knowledge
- easily accessible and findable datasets
- clarity about who is accountable for the datasets
- evidence about the value and impact of other data initiatives within the public sector
- sufficient funding to explore new service delivery models.

We would also benefit from better cross-department/team working, and engagement and buy-in from senior management.

If you could change one thing about data in government, what would it be?

That it is published and used in accordance with the needs of service users.

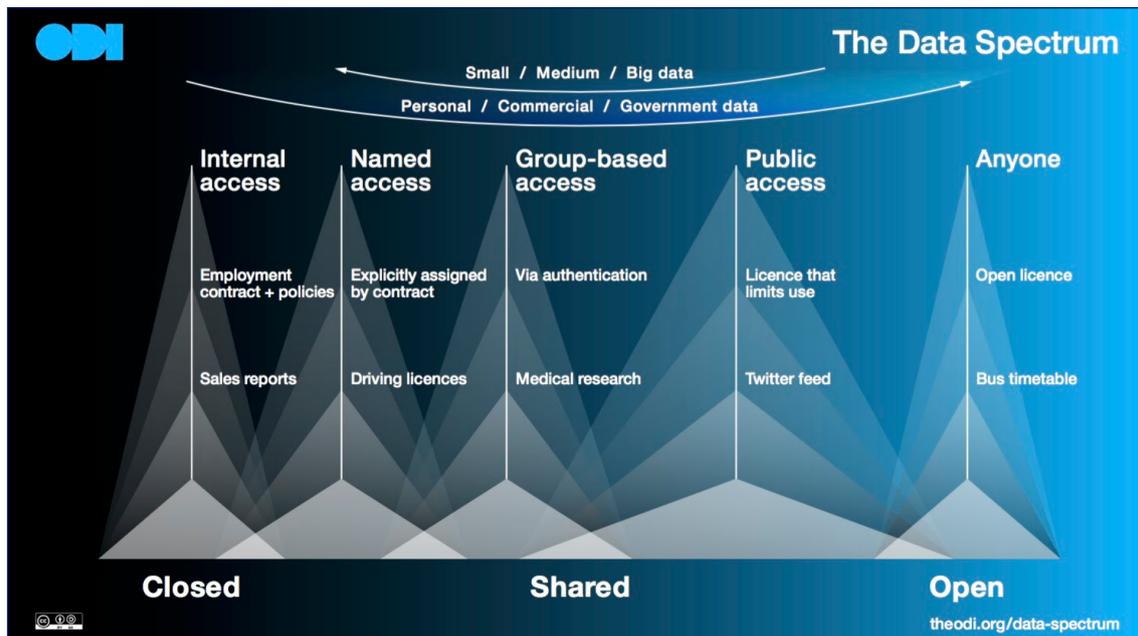
Key links

[Public service design](#) (collection)

[Using Open Data for Public Services](#) (report)

[Data and Public Services Toolkit](#) (resources)

[Making our cities 'open'](#) (project)



Data Bites #8

February 2020

Sponsored by



Presentation	8:1
Topic	How data transparency can be a good way of tackling inequality
Speaker	Marcus Bell, Director of the Race Disparity Unit (RDU)
Organisation	Cabinet Office

Can you summarise your project in a few sentences?

RDU collects, publishes and analyses government data about ethnic disparities in order to inform the public and others and also to support the development of government policy.

What problem are you trying to solve?

Government data about ethnicity is hard to access and interpret. Such data is often contentious and so needs to be presented with care and objectivity.

What difference will it make to citizens?

The data provides information about ethnicity in the UK and how outcomes in key public services are different for people from different ethnic backgrounds.

How will you know when it has succeeded?

When ethnic disparities that cannot be explained by other factors (and this is important) have reduced.

What are the main hurdles?

Government departments do not have a consistent approach to classifying ethnicity.

If you could change one thing about data in government, what would it be?

Policy makers would use and understand their own department's data more systematically.

Key link

[Race Disparity Unit](#) (website)



Presentation **8:2**

Topic **The potential of combined data for early intervention and prevention**

Speaker **Stan Gilmour, Head of Public Protection**

Organisation **Thames Valley Police**

"I would make sure that government departments, and public agencies, shared data as standard."

Can you summarise your project in a few sentences?

This project is scoping the benefits that could be realised through sharing multiagency data to identify risk and protective factors to prevent offending and reoffending, and reduce vulnerability and crime.

What problem are you trying to solve?

The problems of silo working that can hamper mechanisms that could identify how best to work for common goals in a system that can sometimes be in unknowing opposition.

What difference will it make to citizens?

Help and support would get to our most vulnerable citizens at the earliest point of need.

How will you know when it has succeeded?

We will identify a drop in demand for late intervention services.

What are the main hurdles?

Funding, and popular support for something that raises real ethical concerns – which will only be forthcoming when we can answer questions of privacy, bias and wider ethics of resource allocation, etc.

If you could change one thing about data in government, what would it be?

I would make sure that government departments, and public agencies, shared data as standard.



Presentation **8:3**

Topic **The data the government doesn't collect**

Speaker **Anna Powell-Smith, Founder**

Organisation **Missing Numbers**

“This series is about people doing interesting things with data in government. I’m talking about the opposite, which is about people not being able to do interesting things, because of there not being data.”

Missing numbers in government data

A missing number is “important, public data the government does not collect”. Examples include complaints about jobcentres (government publishes complaints about hospitals, and collects them on schools, but does neither for jobcentres); various numbers on land and housing, including land ownership concentration and control, on the rental market, affordable housing, land values; and social care, including the quality of adult social care and meaningful outcomes, the cost and extent of private care, and value and extent of unpaid care (also highlighted by the Institute for Government, Doteveryone and the Office for Statistics Regulation).

It may not be considered important by the people with the power to produce it, it may be embarrassing, it may be difficult to collect, it may be all of the above. The UK statistical system is better at producing accurate data than relevant data – although the latter is a statutory duty, and things are starting to change.

Drawing on the criteria for national statistics, it may be a number used to allocate public resources, on a topic of significant public interest, generating regular media reporting, or in parliamentary questions and ministerial speeches.

Ways of fixing missing numbers might include a Bean review (an independent report into economic statistics), but for social statistics. Focus on who’s using these stats. Identify missing numbers routinely in public life.

Key link

[Missing Numbers](#) (website)



Presentation	8:4
Topic	Improving the use of data across government: how big is the challenge?
Speaker	Dr Stephen Lorimer, Head of Public Sector Data Policy
Organisation	Department for Digital, Culture, Media and Sport (DCMS)

Data at DCMS

DCMS has a job to make data useful, for the services being delivered by government and policy makers. There are lots of great things going on in government, beyond the scope of this presentation: Application Programming Interfaces, technical data standards (GDS has plenty of talks on those), the National Data Strategy, GDPR, work on online targeting, and more.

According to one study, in 2012, only 0.5% of data was analysed. There's also the 80/20 data science dilemma – we spend almost all the time cleaning data, not doing data science. How do we make it useful?

Better decisions. Service design. Evidence. Value: what is the value ecosystem, not just government but its interaction with the wider economy? What are the challenges? Data is a strategic asset, but it is unexhausted and reusable (it's not as if someone has exclusive access to a dataset, exclusively, forever). How government treats it as an asset is very different to, for example, land. Leadership around data is not an easy task.

Do you fund the shiny and new, or the 'fixing the plumbing' jobs (that's been a very successful MHCLG campaign). There are huge challenges around data quality, as shown by the recent [NAO report](#). These include data standards (making sure we're talking about the same data, same person, same objects), legacy IT, data sharing (including understanding of the Digital Economy Act) and working in silos (how do we share the cost and benefits of data extraction?).

Recent [Treasury minutes](#) list some of DCMS's priorities and milestones, including mandating a consistent approach, ways to monitor compliance, and ensuring that sweeping away legacy IT results in more useful data rather than replicating the same problems with a new set of suppliers. They are also keen to hear ideas.



Praise for Data Bites

I really enjoyed Data Bites. Eight minutes is sufficient time to get across what can be possible, and sparked some interesting discussions afterwards, with a diverse group of people. I'm just disappointed that I live too far away to be there every month!

Louisa Nolan, ONS (presenter at DB #1)

It was a highlight of 2019!

Jon Downing, Ofgem (presenter at DB #1)

Presenting at the IfG gave us exposure to a wider government audience in both central and local government and that it enabled us to meet and network with key influencers.

Martin Pilkington, RHE Global (sponsor/presenter at DB #2)

I thought Data Bites was a really good way to reach a diverse audience with people outside our usual world of official statistics. As I'm not based in London I really like the fact it is live-streamed, covered on social media and the talks are available on YouTube. It's an event we would like to participate in and support in future.

Catherine Bromley, Office for Statistics Regulation (sponsor/presenter at DB #4)

Such a great meetup, the whole event has such a brilliant welcoming atmosphere.

Anna Powell-Smith, Missing Numbers (presenter at DB #8)

Data Bites is **the** best public sector event I go to. A big achievement to get such consistently good presentations, a great format, and a great mix of attendees.

Steve Parks, Convivio

I really enjoyed it. The brisk format was excellent... I am by no means a data nerd but was very stimulated by the whole event. Went away buzzing with ideas.

John McTernan, Burson Cohn & Wolfe and former adviser to Tony Blair

About the author

Gavin Freeguard

Gavin joined the Institute for Government in August 2013 and leads the organisation's work on data, transparency and digital government, including our [Whitehall Monitor](#) project. He has also worked on government contracting and preparing politicians for government.

Gavin was previously political adviser on culture, media and sport to Harriet Harman MP and, before that, deputy director of the Orwell Prize and senior researcher at the Media Standards Trust. He holds an undergraduate degree in Modern History and Politics from the University of Oxford (2007) and a Master's degree in the History of Political Thought and Intellectual History from UCL and Queen Mary, University of London (2012).

Gavin is a trustee of [the Orwell Youth Prize](#), a member of the [UK Open Government Network](#) steering group, and a member of the research commissioning board of the ESRC's [Administrative Data Research \(ADR-UK\)](#) project. He represented the Institute on the advisory board to the Treasury's [Government Financial Reporting Review](#). He sends a weekly email on data and data visualisation called [Warning: Graphic Content](#).

Acknowledgements

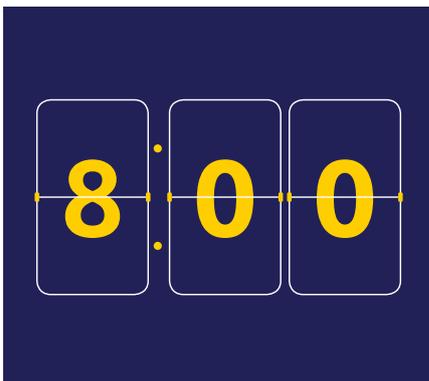
Data Bites would not be possible without the support of many people across the Institute, including Team Data with their ideas and enthusiasm (and live-tweeting), but particular thanks are due to our events and communications team, especially Magnus Summers, Penny Cranford, Melissa Ittoo and Neil Bowerman for making things run as smoothly as they have. Huge thanks to our dynamic partnerships duo, Pritesh Mistry and David Tripepi-Lewis, for their hard work in keeping the series funded. Many thanks, too, to Will Driscoll and Rowena Mayhew for all their work on this report.

Each Data Bites event ends with three sets of thank yous: to our sponsors, to our audience and to our speakers. The success of the series, as well as this report, simply would not have been possible without supportive sponsors and intelligent and insightful audiences, both online and in the building, so thank you again to them. But the biggest thanks must go to our 36 wonderful speakers. I've thoroughly enjoyed listening to and learning from them; those lessons have lasted long past their eight minutes.

The Institute for Government is the leading think tank working to make government more effective.

We provide rigorous research and analysis, topical commentary and public events to explore the key challenges facing government.

We offer a space for discussion and fresh thinking, to help senior politicians and civil servants think differently and bring about change.



 instituteforgovernment.org.uk

 enquiries@instituteforgovernment.org.uk

 +44 (0) 20 7747 0400  +44 (0) 20 7766 0700

 [@instituteforgov](https://twitter.com/instituteforgov)

**Institute for Government, 2 Carlton Gardens
London SW1Y 5AA, United Kingdom**

© Institute for Government 2020

The Institute for Government is a registered charity in England and Wales (No.1123926) with cross-party governance. Our main funder is the Gatsby Charitable Foundation, one of the Sainsbury Family Charitable Trusts.