[Speech: Big data in government: the challenges and opportunities](#)

Ladies and Gentlemen, good morning.

I'd like to begin by thanking Reform for giving me the opportunity to address you today. I want to talk to you about the potential of big data in government — and the hurdles we face along the way.

We often hear that this is the "age of data", or that data is the raw material of a new industrial revolution. There's truth in this. And there's huge opportunity.

Data can truly be a catalyst for a society, an economy, a country that works for everyone.

Of course, data isn't new. There has always been data. The Domesday Book is data. The Rosetta Stone is data.

But the rapid advances in technology and the development of analytical tools and techniques mean we can now gather and share data in huge quantities. We can process and analyse it at previously unimaginable speed. We can draw conclusions and create policies and services that reflect how people live now.

And we can help them live better, more securely, more healthily and more prosperously as a result.

Data entrepreneurs are mining public sector data to create apps and services to make our lives more convenient. Services driven by open data are already giving people more choice in where they get their healthcare, where they live and where their children go to school. There's even a Great British Public Toilet app — a sort of relief map of the country!

In government, we get this. We've always held enormous quantities of data — now we need to make sure we use it properly. Getting this right is the next phase of public service modernisation.

That's why this month we have published the [Government Transformation strategy](#). And the Digital Economy Bill is in the last stage of its journey into law.

There are 3 key areas of opportunity that we need to grasp:

- first, improving the experience of the citizen
- second, making government more efficient
- third, boosting business and the wider economy.

The impact of data analytics and big data in our lives — for example the way online retailers tailor their recommendations for the food, books and music

we buy – is quite familiar.

Less has been said about the transformative power of this technology for the delivery of high-quality public services. And it's time that changed.

With the evidence of data we can spend less time developing policy and services that don't work, and instead focus on continuously improving those that do.

I want people to turn to digital public services as readily and confidently as they do when shopping, socialising or checking bus times.

By doing so, we can actually change the way citizens interact with us – making the relationship we have with them more transparent, more responsive, and based on increasing levels of trust.

For big service delivery departments like the Home Office, HMRC and DWP, data analytics means the ability to search across organisational data sets. It can provide data for operational teams to put into practical effect.

In DWP, for example, providing job seekers with more targeted advice, and opportunities that closely match their personal profiles. The department is also working on data-informed tools, such as interactive visualisations of benefit claimant trends.

There are examples at home and abroad where data is being used to address people's real concerns about their daily lives; providing solutions that were not available before.

In June last year, for example, Land Registry and partners published the first UK House Price Index, and provided a single source of information as opposed to the multiple competing versions which existed before.

Land Registry data has also been used to create a range of information services. From whether rude-sounding street names have an impact on house values – (they do!); to more serious matters, such as whether your home is on a floodplain.

Land Registry's Flood Risk Indicator service uses data from the Environment Agency to identify flood risk for any registered piece of land within England and Wales.

The Companies House Service gives us free access to real-time information on companies. It's receiving millions of search requests every day from people checking supplier and customer information.

The service can also be used for more mundane but practical reasons – if you're getting in builders to do work on your home, you can go on the Companies House website and check them out first.

Healthcare is another exciting area. Moorfields Eye Hospital and DeepMind Health are partners in a research project that could lead to earlier detection of eye diseases.

At the moment, clinicians rely on complex digital eye scans. 3,000 of these scans are made every week at Moorfields. But traditional tools can't explore them fully, and analysis takes time.

Moorfields will share a data set of one million anonymised scans with DeepMind, who will analyse them using machine-learning technology. This can detect and learn patterns from data in seconds, to quickly diagnose whether a condition is urgent.

With sight loss predicted to double by 2050, the use of cutting-edge technology is absolutely vital. The right treatment at the right time can prevent many cases of blindness or partial sightedness. Up to 98% of sight loss resulting from diabetes, for example, can be prevented by early detection and treatment.

Analysing data can also play a direct and powerful role in protecting the most vulnerable in society.

The Home Office Child Abuse Image Database has transformed the investigation of child abuse crimes and child protection. It won the Civil Service Innovation Challenge in 2015.

The database brings together all the images of abuse that police find. Using the images' unique identifiers and metadata, they can check devices they've seized from suspects against the material on the database much more quickly.

Previously a case involving, say, 10,000 images, would typically take up to three days to review. Now, it can be reviewed in an hour.

So, we have a process that is cheaper, less labour-intensive and more efficient. This is all good. And it makes the investigation and prosecution of these appalling crimes vastly more effective.

There are also examples of government data meeting needs that more of us will be familiar with — like tax.

Personal tax accounts from HMRC now take a real-time digital approach. For the first time you can log in when you like, check your tax information and manage your details online in one place. More than 8 million citizens have now signed up, including some of you here today, I expect. HMRC's digital team now has around 30 new online services in development.

Government open data, combined with digital technology, can also fuel an open economy. It will provide information that entrepreneurs, data start-ups and the general public can use.

In 2015, the digital sector contributed £118 billion to the economy, supporting over 1.4 million jobs.

The UK Government was an early world leader in open data. So far, we've released over 30,000 non-personal data sets in machine-readable formats, for no cost, and open for anyone to use or build upon. This has enabled the creation of innovative products that deliver value for citizens.

So far, these data releases have been turned into over 400 different apps. You may well have used some of them yourselves:

- the Floodalerts API: which uses Environment Agency data to provide 15-minute updates about flood risk
- UK Food Hygiene: which lets you see take-away and restaurant food hygiene ratings to help you make decisions on where to eat out
- FillThatHole: a site for reporting potholes and other road hazards across the UK using ONS Census geography data
- There are also apps for finding the best dentists, GPs, schools and universities.

The list of sectors tapping into Defra group data from Lidar (the airborne, laser equivalent of radar) is truly remarkable. British wine producers are using the terrain-mapping data to help them decide where best to plant vines, and if the recent prominence of English sparkling wine is anything to go by — they are having great success!

Architects are using it to build a model of London as they plan the next high-rise building; computer game developers to build new landscapes for Minecraft; and archaeologists to discover lost networks of Roman roads. In October last year alone there were almost 21,000 downloads of Lidar data from data.gov.uk.

Some companies don't only use open public data to build a business but also to act as a positive disruptive force. FoodTrade, for example, maps the food supply chain system, making it easier for people to buy and sell fresh local produce.

And other start-ups are using open data in ways that boost the economy by providing data analysis tools and data products that support the growth of SMEs.

A firm called GeoLytix offers a range of products based on geospatial data — giving smaller companies access to information that they would not be able to do on their own, and helping them to solve business location issues in the process.

So as we look to improve the availability, quality and use of government data as the basis for fully transformed public services, it will also provide a new stimulus for data-based businesses.

Because government data is public data we have a duty to use it well and open it up where possible — and we have to be seen to do so cost-effectively, efficiently, proportionately and appropriately.

But it is not without challenges, and I want to address two in particular:

1. Winning and retaining public confidence
2. Building Civil Service capability in how we collect, store, analyse, share and use data

Public trust is absolutely critical to achieving our ambition for a data-

driven government.

Information and data is power. Which is why, historically, the ability to communicate and understand it was so jealously guarded. Now that we are openly releasing information, we have to do so responsibly.

Trust means giving people confidence that their data is used appropriately and effectively, and that it's secure, particularly when it's being shared by different authorities. That trust has to be earned.

In partnership with civil society, GDS has published an ethical framework for data science in government. It is based on the key principles of data security, openness, user need and public benefit. And it highlights the importance of ensuring the data and models we are using are robust.

And to complement this, the Office for National Statistics has adopted a framework called 'The Five Safes' for building and maintaining trust and confidence:

Safe people Safe projects Safe settings Safe data Safe output

And we are looking outside of government too. The Royal Society and British Academy are conducting an independent investigation into how data is, and could be, used by government, and the types of governance that may be required.

We need to make sure we take the public with us on this journey, and maintain their trust that we are using and sharing data responsibly and effectively.

Transparency is part of this — transparency of evidence, 'showing your working', and opening up to greater scrutiny the data and analysis on which we base policy decisions. For transactions (such as driving licence and passport applications) users can now see the data government holds about them and change it if it is wrong.

We must also have the confidence that a person accessing a service is who they say they are, and we must do that in a way which the public trust.

Verify — the government identity service for citizens — is enabling people to access a whole range of online government services easily, securely and in a way which builds their trust. By 2020, we are aiming to get 25 million people using the service.

But providing transactional services is only part of what government does — it also uses data to: identify individuals for support where there is wider impact on society — such as elderly people in fuel poverty; or to develop policy, plan services and assess outcomes; or to promote innovation, or allow citizens to hold us to account.

The introduction of new legislation on data access in the Digital Economy Bill is designed to give confidence that government is doing the right thing.

The Bill provides a robust legal framework for sharing data between public

authorities, where there is a clear public need and benefit.

A case in point is the Troubled Families Programme to get children back into school, put adults in employment or on a path back to work, and cut youth crime and anti-social behaviour. To identify families in need of help, public authorities need to see information held by other authorities.

The Bill also makes provision for public authorities sharing information with energy companies to identify customers living in fuel poverty so they can automatically receive support — such as energy bill rebates or energy-saving measures.

We see a clear link between public trust and government capability in its handling of data.

A May 2016 survey for the Government Data Science Partnership showed that public approval for government sharing data is actually quite high when they accept that it is used in measured, proportionate and targeted ways. So the way we collect, store and release our data must keep pace with expectations.

We've introduced the first six developer-friendly open registers of reliable and up-to-date data on specific areas that we can use with confidence to build a service. These include countries, territories and English local authority registers — with more to come, covering everything from police stations and schools, to doctors and courts.

We need to think about the collection and storage of data as part of core national infrastructure, in the same way as we think of our road and rail systems, energy supply, and telecommunications networks as infrastructure.

To give confidence that we handle data and can realise its potential effectively, we also need the right people with the right skills in the right place, in government and across the economy.

And we are still short of the key data science skills in Government.

Such a shortage is not peculiar to this country — research in the US predicts that by 2018 they will be short of 190,000 data scientists.

Here, in the UK Civil Service, we are growing the specialist data science community in a variety of ways — from direct recruitment to training to defining new career pathways for analysts.

The Data Science Accelerator Programme is tapping into the 3,000 or so analysts from other disciplines looking to develop their data science skills

A Data Science Campus opened its doors at ONS's headquarters in Newport last October. And the first intake for a new Apprenticeship in Data Analytics started work on their two-year vocational training programme at the end of 2016.

And because everyone at every level should have an appreciation of the power of data, we're developing a programme in data literacy for non-data

specialists. The Digital Academy will provide skills training right across government for up to 3,000 people a year.

Together, these measures are nudging us towards a cultural shift in the status of data in government and those who work with it.

And how Government uses data in service of the citizen will define how the citizen experiences Government. When we get it right, we will deliver the right service at the right time to the right person. And that is our goal.

So — to conclude.

Data underpins everything we do — but we could do so much more. The possibilities are tremendously exciting: Services for the the citizen that are both targeted and responsive; More effective and more efficient government; and Data as an enabler of growth in the wider economy.

And we have to step up our efforts. We have announced that we plan to appoint a new Chief Data Officer, whose role will be to oversee this agenda, and a cross-government senior Data Advisory Board.

But it is not without challenge, and we need your help. We need to partner with you to help us navigate the difficult ethical judgements about how to share data in the right way.

We need you to tell us which data you want in the open and how, and we need to share scarce resources to build toward our overall goal.

Data is at the heart of 21st century government. It puts the citizen front and centre in public service delivery. It powers effective decision making on the front line. It makes government work for everyone, by better reflecting the world that we live in.

We are at the start of this journey — but I can't wait to see how we can accelerate from here toward a public sector which is truly in service of the citizen.