DIGI GOV EXPO '24

CITIZEN EXPERIENCE THEATRE

THEATRE SPONSORED BY Adobe

The Integrated Data Service: Showcasing the art of the possible

Fiona James

Chief Data Officer and Director Data Growth and Operations

Office for National Statistics (ONS)

Integrated Data Service

Welcome to the art of the possible: Showcasing the IDS in practice

HM Government

In partnership with

Office for National Statistics

Fiona James
Director of Data Growth & Operations and Chief Data Officer,
Office for National Statistics

After today's session you will understand:

Integrated Data Service

How data sharing across government **leads to better and more effective policy decisions -** case studies and use cases

The **'art of the possible' opportunity** to enable faster and wider collaboration

What **we have learned from the IDS** so far to inform future **data driven government transformation**

Integrated Data Service

Data Sharing Challenges

1. fragmented data that is **siloed**
2. **trusting** how secure the process is
3. **resource availability** to access data (often it's a manual process)
4. limited **cross-referencing** capability
5. **data standards** constraints
6. inconsistent **data quality**
7. **inadequate granularity** to address local issues
8. Inefficient, inconsistent and **costly legacy systems**
9. limited **predicative capability**
10. data is not reusable

Integrated Data Service

How data sharing across government leads to better and more effective policy decisions

The IDS brings together ready-to-use data to enable faster and wider collaborative analysis for the public good

The IDS enables rapid, timely and enhanced policy decision making across government, academia, public policy, the commercial sector, and the devolved governments of the UK.

The IDS is for government users, and users from the UK's research community. It will deliver on the government's ambition of **transforming for a digital future**

Enabling first-class decision making

Integrated Data Service

Recent events such as conflict, the pandemic and the cost-of-living crisis show us why **accurate and timely data is critical for decision makers**.

The IDS is providing the **tools and capability to unlock the potential of data to improve the lives of UK citizens.**

To date, sharing data across government remains a challenge, the IDS **helping socialise and break down barriers, enabling data to be more available for more effective policy making**.

IDS is being utilised across govt to unlock improvements in people's lives

Integrated Data Service

More efficient and effective public services

Improved **policies** for society and the economy

Improvement: Concentrate resources on the **most effective treatments** for helping people stay in / return to work

Insight: Identify most effective treatments using **linked health and labour market** data

Improvement: Enhance mechanisms for **devolving power** to improve local services

Insight: Identify areas of most rapid improvements using **linked mobility, labour market and business data**

Improved **operations** for society and the economy

Improvement: Increase retention and reemployment of **trained nurses and midwives**

Insight: Identify career patterns of registered nurses and midwives using **linked profession registration and labour market** data

Improvement: Prevent **inequalities** in the operation of the tax system

Insight: Identify unintended inequalities using **linked tax customer data and Census 2021 demographic data**

More productive economy

Improved **regulation/ opportunity** for labour market productivity

Improvement: Greater **job opportunities** through better support of local labour markets

Insight: Identify dynamics of local labour markets using **linked labour market and firm level** data

Improvement: Raise productivity and health by promoting **workplace best practice**

Insight: Identify most effective practices using **linked firm productivity and mental health data and Census 2021 demographic data**

Better society

Improved **living standards for everyone in society**

Improvement: Improve **housing standards** so people can live healthier lives

Insight: Identify housing conditions most associated with ill health using **linked housing quality and health data**

Improvement: Make people safer reducing prevalence of **crime hotspots** created by gatherings of people

Insight: Identify occurrence and nature of transitory crime hotspots using **linked mobility and crime data**

Integrated Data Service

Case Studies and Use Case

Integrated Data Service

Case study:

Connecting data and transforming public policy
A case study on health and labour market analysis

Connecting data and transforming public policy
A case study on health and labour market analysis

Integrated Data Service

Challenges to solve

- manual data linkage was time-consuming
- resource-intensive
- high operational costs
- negatively impacting quality
- speed of analysis
- linking datasets was complex

The need

Understand the connection between chronic health conditions and inactivity in the labour market

Data shared and linked

By a demographic index:

- Hospital Episodes
- Labour Force Survey
- General Practitioner

Aim Develop targeted policy interventions to support individuals in returning to work, while informing future budgets

HM Revenue & Customs | Office for National Statistics | HM Treasury | NHS

Connecting data and transforming public policy
A case study on health and labour market analysis

Integrated Data Service

Outcomes

- overcame barriers for accessing and matching challenging datasets
- established methods to join datasets
- provided flexibility through variables
- delivered efficient and thorough data analysis with reduced cost and resourcing
- successfully analysed different datasets to uncover hidden patterns
- provided a more complete picture of health issues

Benefits

Increased efficiency

Departments were able to quickly and easily gather, analyse and compare data, for projects at a faster pace, while reducing time and resource costs

Improved effectiveness

By standardising and linking data, datasets could be compared to make more evidence based, targeted policy intervention decisions

HM Revenue & Customs | Office for National Statistics | HM Treasury | NHS

Integrated Data Service

Case Study:

Reducing the cost of commercial data for better decision making

Reducing the cost of commercial data for better decision making

Integrated Data Service

The need

Understand employment rates and potential job opportunities by simplifying how government use commercial data sourced from the private sector

Data shared

- job vacancy data
- open source geography data
- O_2 Telecoms Data

Challenges

No centralised process for getting commercial data resulting in an inconsistent, fragmented and expensive approach when engaging with private sector data suppliers.

Resource-intensive manual negotiation of separate contracts with each data supplier, causing delays to data access, increasing administrative burdens and complicating data sharing across government departments.

3 projects include:

Office for National Statistics

textkernel

O_2

Department for Levelling Up, Housing & Communities

Department for Education

Case Study: Reducing the cost of commercial data for better decision making

Integrated Data Service

Solution

- contracts negotiated for onward sharing
- a single point of contact in government for commercial data suppliers
- mobile phone data was purchased from O_2 standardised, and linked

Outcomes

- successful data with reduced fees agreed
- for onward cross-departmental sharing
- reduced procurement costs through centrally negotiated data sharing agreements
- real-time understanding of population numbers through new data

Benefits include

efficient

effective

reduced costs

Integrated Data Service

Use case:

Combating £9.4bn of fraud and error through data sharing

Combating £9-10bn of fraud and error through data sharing

Department for Work & Pensions

Integrated Data Service

Limitations

- fragmented data sources
- delays leading to slow response and discrepancies
- inadequate OGD cross-referencing capability
- limited predictive analytics
- privacy concerns
- varying data governance standards

The need

DWP experience significant levels of fraud /error

They must:

- achieve savings of £1.3 billion 2023-24 (fiscal)
- With a cumulative £9.4-£9.7m billion over the next five years
- enhance detecting, preventing and rectifying fraudulent claims and errors, at the earliest possible stage

Guardian headline

“DWP errors leave more than 200,000 pensioners £1.3 billion out of pocket”

Combating £9.4bn of fraud and error through data sharing

Department for Work & Pensions

Integrated Data Service

How the IDS could help

- real-time data integration
- provide a holistic view of claimants' circumstances
- analytics to detect patterns indicative of fraud/error
- secure and ethical data sharing
- easier and efficient access to standardised, ready-to-use data

Potential benefits

For the public

- Transparency
- Understanding
- Resilient system that can adapt

For DWP

- financial savings, reduced fraud and error
  efficient allocation of resources
- allows for better genuine claimant support
- strengthened data security
- enhanced inter-departmental collaboration

Integrated Data Service

IDS Features and Benefits

Integrated Data Service

The ‘art of the possible’ opportunity to enable faster and wider collaboration

Integrated Data Service

Technology

- Cloud-native technology
- Security
- Tooling
- Enables federated data access

Data

- Automated data operations
- Enterprise data agreements
- Data Linkage through a consistent framework
- Indexed data at scale
- Government enterprise data sharing model

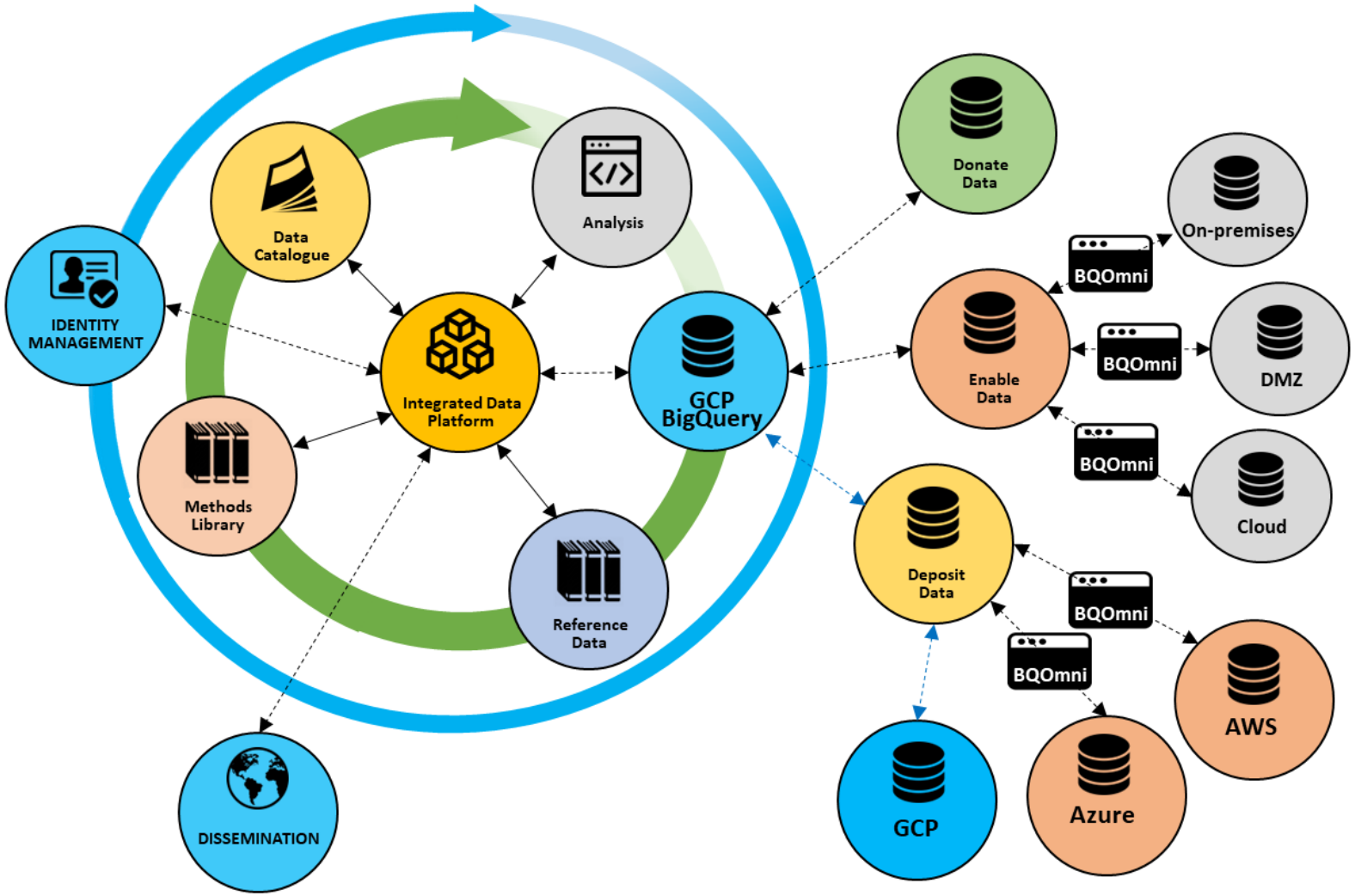
Service

- Statistics and Registration Service Act
- Already Digital Economy Act accredited
- Streamlined processes for government users
- User support

IDS: delivering transformational change

Holistic Technology Architecture

Integrated Data Service



IDS: delivering transformational change

Security

Integrated Data Service

Cloudflare

provides an isolated browser which protects users from accessing untrusted, potentially malicious websites and applications.

JFrog Artifactory

used to house, manage and distribute commonly used artifacts and packages to users.

GitHub

is an internal code hosting platform for version control, collaboration and code-sharing.

JFrog X-Ray

assesses whether packages are considered safe or unsafe before users are able to download into the platform.

GitGuardian

utilised to monitor code repositories.

IDS: delivering transformational change

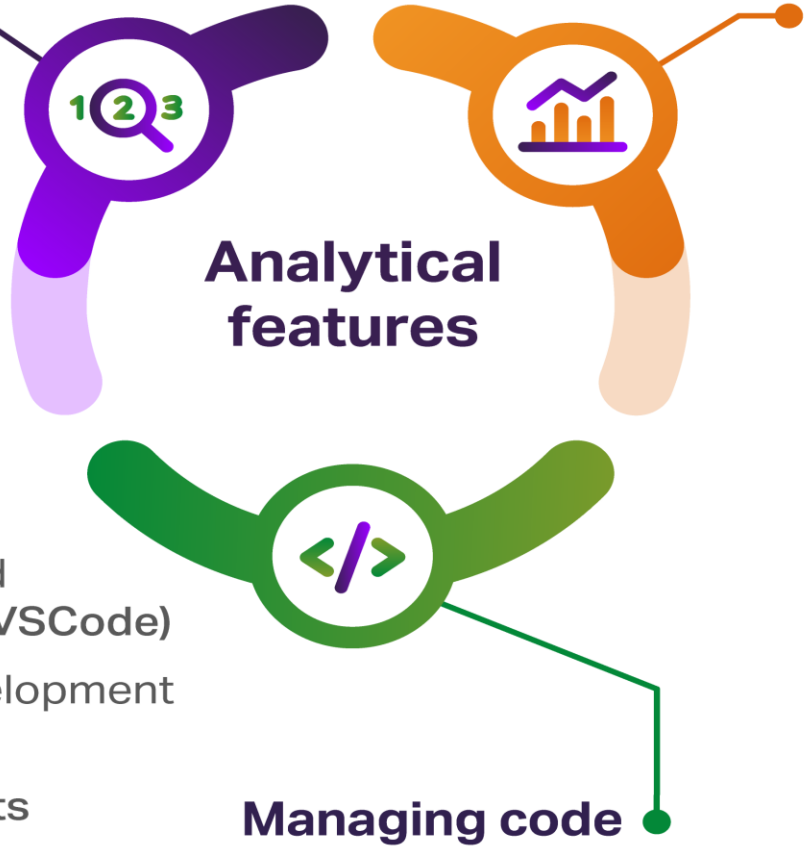
Tooling

Undertaking analysis

- ✔ **Google Vertex Notebooks** (based on Jupyter) to write **Python** and **R**
- ✔ **Google BigQuery** to query and manipulate data using Structured Query Language (SQL)

Coming soon

- Python coding via Integrated Development Environment **(VSCode)**
- R coding via Integrated Development Environment **(R-Studio)**
- Analysis using **Google Sheets** (and plug-ins)
- Large scale analysis via Google Dataproc **(Spark)**



Presenting outputs

- ✔ Coding functionality via **R** and **Python**

Coming soon

- Data visualisation via **Looker**
- Geospatial analysis via **Google BigQuery** geospatial capabilities

Managing code

- ✔ Source code via **Git**
- ✔ Packages via **Artifactory**

Integrated Data Service

IDS: delivering transformational change

Data Access

Integrated Data Service

Anticipating Requirements: Thematic Policy Areas & Integrated Data Products; Essential Shared Data Assets; Commercial data

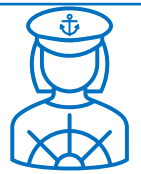
Outcome: *Data acquired proactively in advance of acute user need, removing friction from a reactive approach acquisition*

Broader and simplified data sharing agreements: Data Processing/Sharing Framework Agreements & Donate annexes

Outcome: *Data sharing via IDS agreed at enterprise-level, on behalf of all data owners within organisation, consistent approach across HMG with new data easily added following agreement*

Common Linkage and data views: upfront permission to index by default

***Outcome**: all data indexed and joinable to other datasets, reducing friction and overheads to deliver bespoke linkage for bespoke usages; application of data views as opposed to manual extracts*



Streamlined project approvals: upfront project approvals; Enterprise and Representative Stewardship; programmatic accreditation

***Outcome**: minimising decision points and time to take decisions on usage of data, maximising flexibility to undertake broader and iterative analytical projects under the DEA.*

Trailblazing linkage with the IDS

The Integrated Data Service (IDS) has a Reference Data Management Framework (RDMF), allowing datasets to be linked with pre-allocated indexes.

RDMF

5 reference data indexes, providing matching services.
Datasets are joined with a common ID:

- **Demographic**
- **Business**
- **Classification**
- **Address**
- **Geography**

Having these indexes will reduce the amount of data linkage performed on a case-by-case basis. By indexing and matching individual datasets, it allows for easy and flexible linking at the point of use.

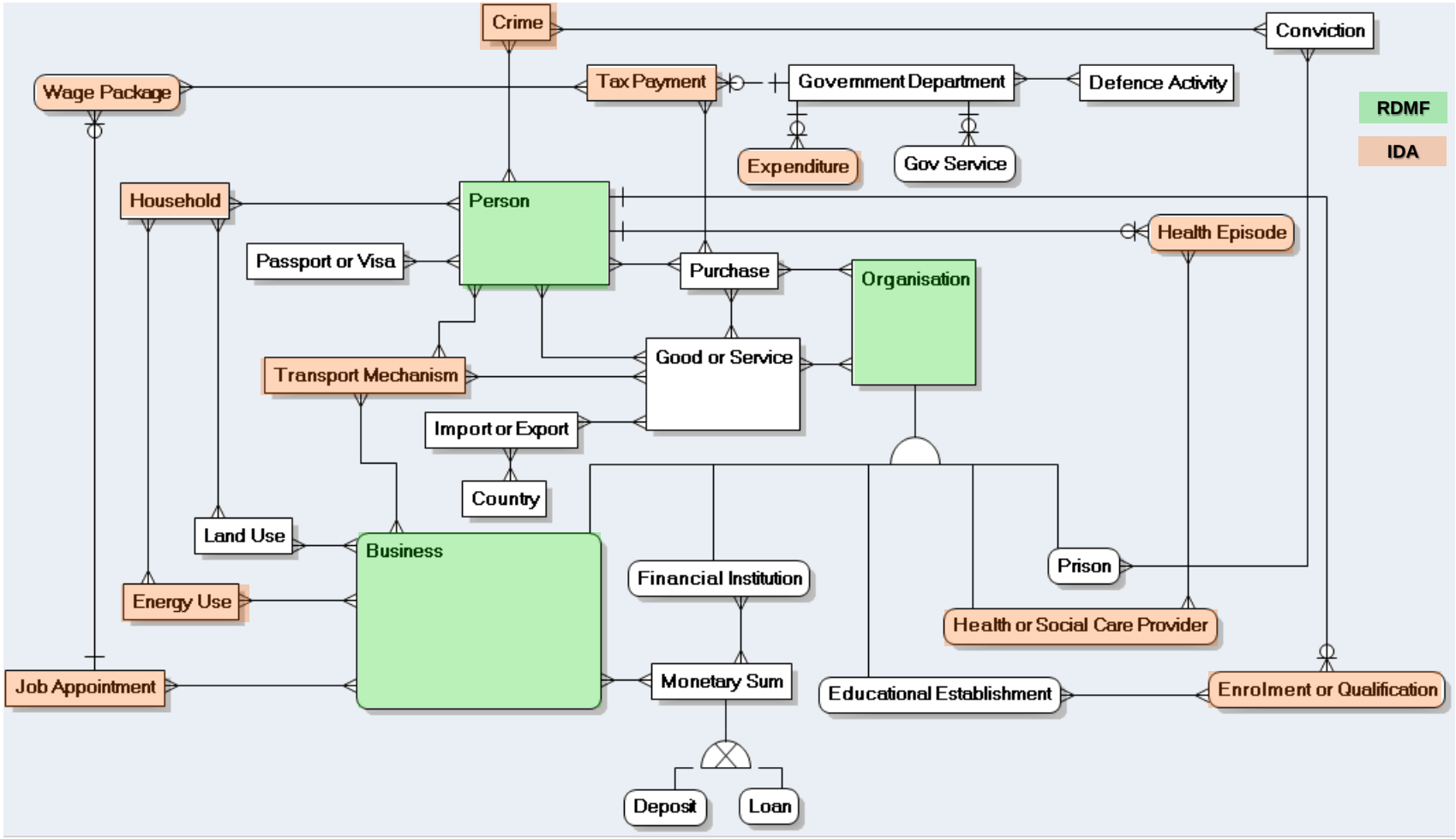
The benefits

It allows:

- Consistency in the same approach to data integration, and **something that potentially all government departments can aspire to through the IDS**
- De-identification of records to facilitate confidential linkage
- Matching services enabled by matching algorithm development that supports faster and more accurate matching
- Quality reporting to assist quality decision making
- Data matching just once, rather than multiple times

The enterprise data model provides much greater scope for cross-cutting analysis and usage

Integrated Data Service



Data Pipeline

Integrated Data Service

The old way – clunky and time consuming

A new policy question is identified

Analysts are engaged

Data requirement identified

Data Acquisition Begins

Data usage negotiated

Data engineered and linked

Data are available

Analysis

Outputs disseminated

Project closes

The IDS is transforming – faster and efficient

Data assets are anonymised, indexed, and linked

A new policy question is identified

Data integration and analysis

Outputs disseminated and semi-automated

Futureproofing: Faster and richer linked data, answering the questions we don't yet know will be asked

Integrated Data Service

What we have learned from the IDS so far to inform future data driven government transformation

Progress on Transforming The Linked Data Infrastructure

Integrated Data Service

Gaining permission to prepare data

Problem: ad hoc data shares using multiple legal gateways take months / years to negotiate

Transformation: **systemic data sharing** agreements under single legal gateway negotiated once then used for repeated purposes

State of play: **slow but making progress**

Indexing and linking data

Problem: ad hoc linkage between datasets using multiple different methods take time and resource for each new linkage

Transformation: systemic **indexing to reference data** allowing multiple subsequent linkages (index once, link often)

State of play: **key tech built and in use**

Finding & accessing linked data

Problem: analysts have difficulty finding information about linked data and it takes months to access it once found

Transformation: comprehensive, **scalable government data asset** with standardised catalogue and **access to data** within days

State of play: **high value government data asset** growing, with indexing

Utilising data within policy processes

Problem: limited compute power and secure environment prevents rapid iterations necessary within policy processes

Transformation: **cloud-based analytic infrastructure** optimised for government use allowing cross-organisational working

State of play: **operational but on-going issues balancing security with usability**

Three key observations from implementation:

Integrated Data Service

1. There is a difficult tension between usability and ease of data availability alongside keeping personal data secure
2. Departments have limited skills, resourcing and bandwidth to engineer data to standards outside of their primary departmental need
3. The systems and services across government reflect clear lines between operational or research use, despite data and linkage operations having similarities

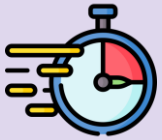
Potential for automation is high, but data needs a lot of work upfront

Integrated Data Service

The first table was loaded into the IDS on the 1st December 2022.

Since then, we have successfully loaded

Total Tables: **6,867** + Total Datasets: **116 (May)** + Total rows of data processed: ~**83 Billion**



This is **200%** increase in data from November to May 2024
with **1,750** tables being created in the busiest single week!

First fully automated regularly updated dataset and data feed into IDS
O2 mobility data

<u>Largest Single Load Job</u>

HES – Hospital Episode Statistics

Total Big Query Storage Usage **2.2 TB**
Load runtime: **17:16 minutes**

Integrated Data Service

Future Vision and Conclusion

‘Art of the possible’ for system-wide data sharing

Government Data

Credit Card

Mobile Phone

NHS Data

IDS

One Log In

Essential Shared Data

Data Marketplace

Operational Services/Systems

Kickstart economic growth

How do we invest in key skills in communities?

Improve living standards

How can we bring down living costs?

NHS Fit for the Future

What are the outcomes for people who use mental health services?

Clean Energy Superpower

Future working patterns that will impact transport and domestic energy emission?

Common Activities

- Data standards & quality
- Data sharing governance & practice
- Data & technical architecture
- Public trust
- Capability
- Infrastructure
- Legislation

Crisis response

Turning 18

Starting a business

Changing jobs

Having a baby

Insight driving Government policy and interventions

Better use of Government Data

Joined up and personalised public services

Please stay connected with us

Integrated Data Service

Sign-up to our newsletter

Scan or email: ids.comms@ons.gov.uk

Learn more

Scan or visit our IDS website:
www.integrateddataservice.gov.uk

With the IDS, the possibilities are endless.

Thank you

Integrated
Data Service

Questions: