The HMRC Datalab: Sharing administrative and survey data on taxation with the research community

Daniele Bega
HMRC Datalab
Facing The Future Conference
22nd November 2013

Aims of the Presentation

- Overview of Her Majesty's Revenue and Customs

- Background on the legal context in which the Datalab and HMRC operate

- Information collected and how it can be accessed

- The HMRC Datalab: details on governance, data available and how it works in practice

1) Her Majesty's Revenue and Customs (HMRC)

Who are we?

- Her Majesty's Revenue and Customs (HMRC) is the UK's tax authority

- Created by Act of Parliament in 2005

- Non-ministerial Department

- Reporting to Parliament through our Treasury Minister who oversees our spending

- Working in partnership with Her Majesty's Treasury

HMRC's role

We **collect** and **administer**:

Direct taxes – paid by you or your business on money you earn or capital you gain.

- Capital Gains Tax
- Corporation Tax
- Income Tax
- Inheritance Tax
- National Insurance Contributions

Indirect taxes - paid by you or your business on money spent on goods or services.

- Excise duties
- Insurance Premium Tax
- Petroleum Revenue Tax
- Stamp Duty
- Stamp Duty Land Tax
- Stamp Duty Reserve Tax
- VAT

HM Revenue & Customs

HMRC's role (continued)

We pay and administer:

- Child Benefit
- Child Trust Fund
- Tax Credit

We enforce and administer:

- Environmental taxes
- National Minimum Wage enforcement
- Recovery of student loans
- Customs duties including collection of data on imports and exports

We are the parent department to the Valuation Office Agency (VOA)
and its commercial arm – the District Valuer Services (DVS).

We also work with the Adjudicator's Office.

HM Revenue
& Customs

2) Our legislation

HMRC's Legislation

- HMRC was created by the Commissioners for Revenue and Customs Act 2005 (CRCA), merging the Inland Revenue and Customs and Excise.

- Section 18 CRCA prohibits the disclosure of any information held by HMRC in connection with its functions to any person, unless there is a lawful authority to do so.

- This includes disclosure to other Government Departments and their agencies, local authorities, the police or any other public bodies.

- However, there are some exceptions that allow lawful disclosure

Exceptions

Where there is UK or EU legislation that permits disclosure ('legislative gateway')

With the consent of the subject(s) of the information

Where the disclosure is made for the purpose of an HMRC function ('functions gateway')
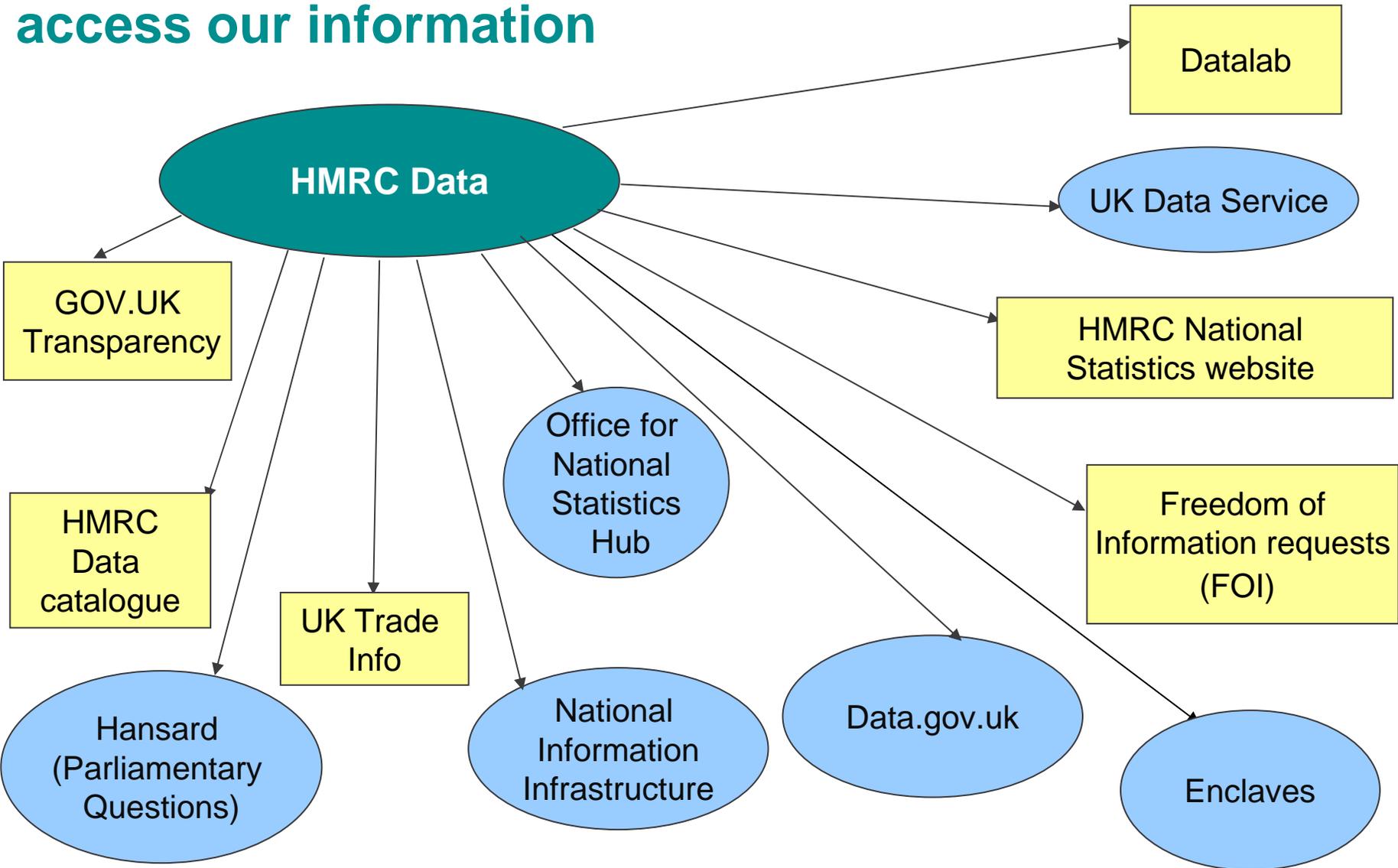
HMRC's functions

- Powers and duties set out in CRCA (or in other legislation), primarily the:
    - Assessment and collection of tax and
    - Payment and management of tax credits.

- HMRC's ancillary functions include:
    - Promoting publicity about the tax system;
    - Establishing advisory bodies;
    - Entering into agreements; and
    - Acquiring and disposing of property.

Disclosure for an ancillary function is permitted where there is a sufficiently close connection between the purpose for disclosure and a core HMRC function.

3) Data Sharing

How the general public can currently access our information

HMRC Data

- Datalab
- UK Data Service
- GOV.UK Transparency
- HMRC National Statistics website
- HMRC Data catalogue
- UK Trade Info
- Office for National Statistics Hub
- Freedom of Information requests (FOI)
- Hansard (Parliamentary Questions)
- National Information Infrastructure
- Data.gov.uk
- Enclaves

HM Revenue & Customs

Legal gateways

HMRC shares information through about 250 information gateways large number of third parties, including:

- Other Government Departments
- Agencies
- Devolved Governments
- Local Authorities
- Other Countries (double taxation agreements)
- The Bank of England

The terms of each information gateway are specific as to the type of information that can be disclosed and its purposes

Public consultation

• In July 2013, HMRC launched a consultation 'Sharing and Publishing Data for the Public Benefit': https://www.gov.uk/government/consultations/sharing-and-publishing-data-for-public-benefit

Various questions, including:
Do you agree that the legal constraints on sharing general and aggregate information should be removed, on a permissive basis, in respect of:
(i) HMRC making information generally available through publishing?
(ii) HMRC sharing information with specific third parties to deliver public benefits wider than HMRC's functions?

Do you agree that HMRC should be able to share anonymised individual level data for the purposes of research and analysis to deliver public benefits wider than HMRC's own functions?

• Results currently being analysed

HM Revenue & Customs

4) The HMRC Datalab

What is the Datalab?

The idea

A **secure** environment where researchers can access, free of charge, **anonymised** taxpayer and customs data to undertake research that serves one of HMRC's functions and benefits the wider research community

The benefits

Improved evidence base and transparency with appropriate safeguards

At the moment the Datalab is only available to UK based institutions, Although we allow collaborations with overseas organisations

HM Revenue & Customs

History of the Datalab

2006 — Initial idea of the Datalab forms, following the example of ONS

2007 — Proposal from Oxford University Centre for Business Taxation (OUCBT) to access corporation tax return data

Statistics Commission Report No. 37: Tax Records as a Statistical Resource: A Review, November 2007

2008 — Feasibility assessment and discussions

2009 — Initial governance designed

2010 — UK Government's Transparency agenda launched

Oxford University Centre for Business Taxation Datalab pilot study

2011 —

The Datalab is opened in May 2011

HM Revenue & Customs

Mining leviathan

In July 2013, the Economist published an article about the Datalab

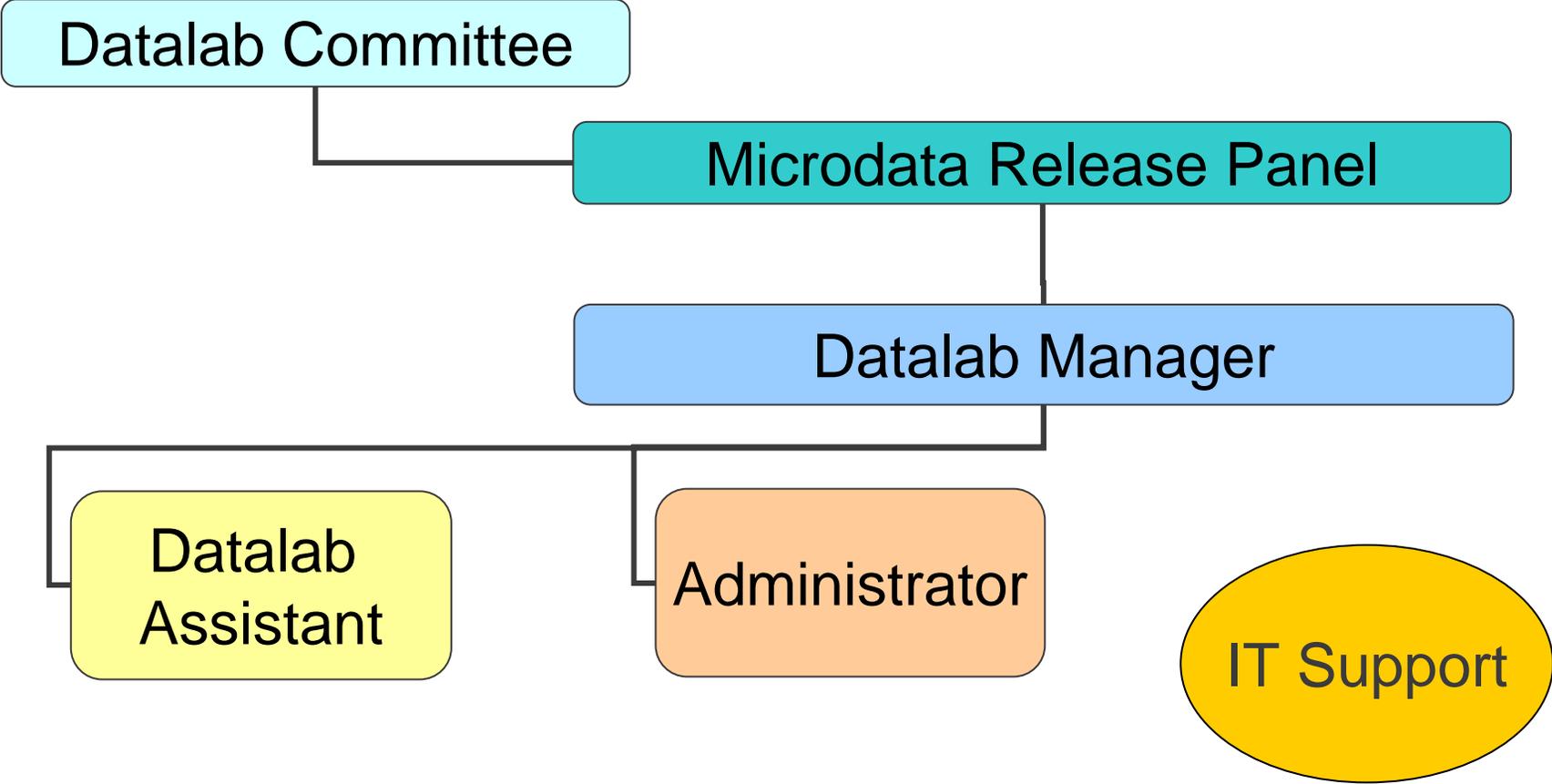http://www.economist.com/news/britain/21582011-surprising-findings-new-stash-government-tax-data-mining-leviathan

'The lab is splendid, but HMRC is rather fussy about its data'
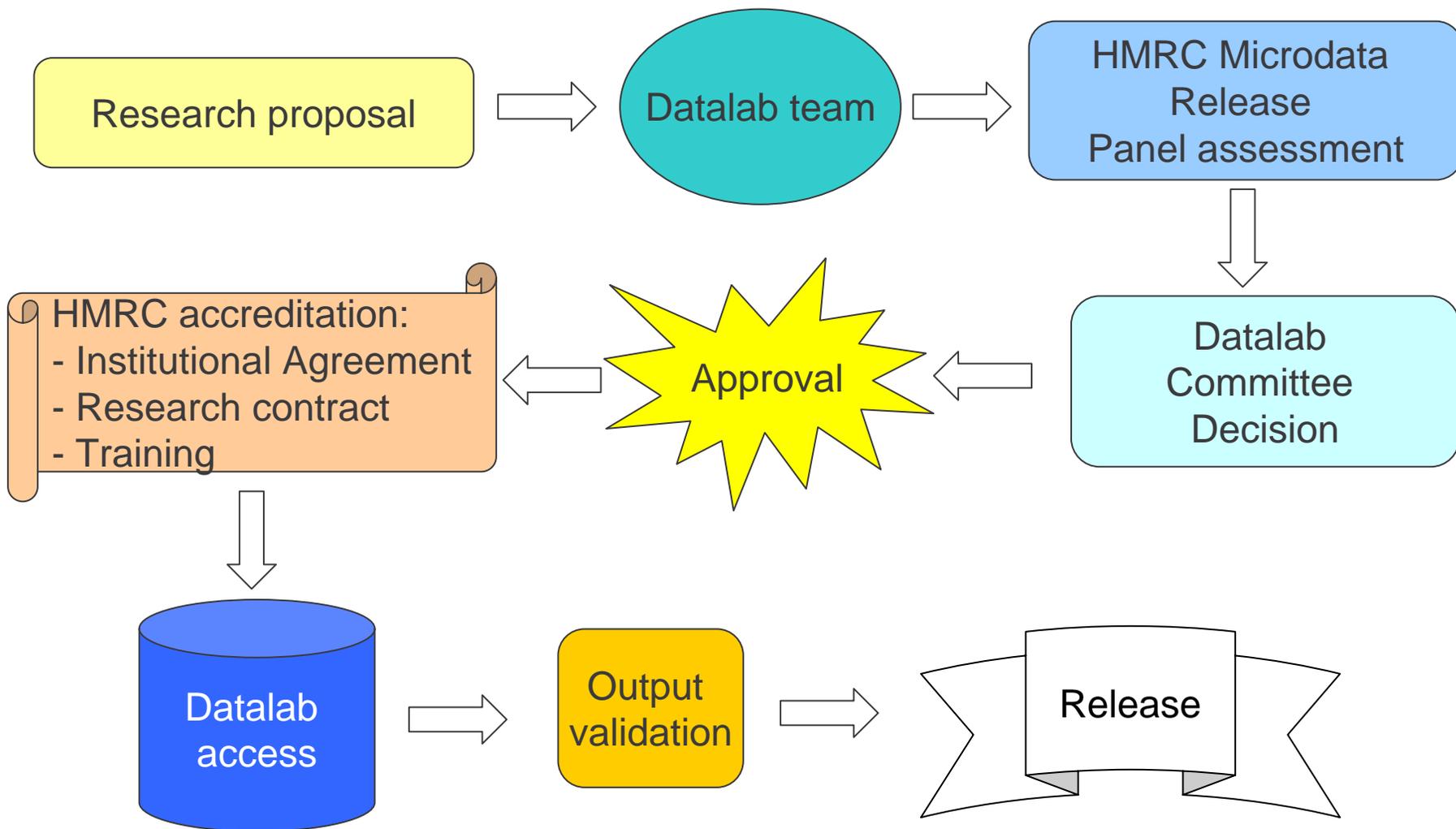
How does the Datalab work?

Following the principles of the ONS Virtual Microdata Laboratory (VML) and Secure Data Service (SDS):

- **Safe projects** - Valid research purpose of interest to HMRC

- **Safe people** - Trusted researchers from trusted institutions

- **Safe data** - Anonymisation of data

- **Safe settings** - Controlled IT environment

- **Safe outputs** - Disclosure controls of any output

HM Revenue & Customs

Governance of the Datalab

Datalab Committee

Microdata Release Panel

Datalab Manager

Datalab Assistant

Administrator

IT Support

Current model for Datalab Access and Output Release

Research proposal → Datalab team → HMRC Microdata Release Panel assessment

HMRC accreditation:
- Institutional Agreement
- Research contract
- Training

← Approval ← Datalab Committee Decision

Datalab access → Output validation → Release

HM Revenue & Customs

Data available

Current datasets:

Compliance Perceptions Survey

Corporation Tax

HMRC Customer Survey

PAYE

Self Assessment

Stamp Duty Land Tax

Survey of Personal Incomes (Public Use Tapes)

Tax Credits

Trade Statistics

Value Added Tax

HM Revenue & Customs

The Datalab environment

- 8 computers - (64-bit PCs, 2 with 24GB and 6 with 32GB RAM)

- Multi Terabyte Server Capacity

- Software including:

  StataMP 12

  SAS 9.3

  PASW Modeler 13 (Clementine)

  MS SQL Server 2005 & 2008

  MS Office 2003 and 2007 inc. Access, Excel, Word & Powerpoint

  Data Matching tools e.g. DQ Global (Match)

Some institutions who have applied to use the Datalab

- Oxford University

- London School of Economics

- Nottingham University

- Imperial College

- Warwick University

- IFS

- Essex University

26 projects approved and about 40 researchers accredited so far

Examples of papers produced to date:

- Corporation Tax in the United Kingdom

- The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records

- Housing Market Responses to Transaction Taxes: Evidence from Notches and Stimulus in the UK

- Optimization Frictions in the Choice of UK Flat Rate Scheme of VAT

- The Investment Effect of Taxation: Evidence from a Corporate Tax Kink

HM Revenue & Customs

Accreditation of researchers

• In order to use the Datalab, researchers need to undertake a training course, consisting of four modules:

        - Introduction to the Datalab
        - Keeping data safe
        - Statistical disclosure control
        - Bookings and outputs

•The training course is valid for 2 years

• In addition, researchers and their institution sign an agreement with HMRC, highlighting provide arrangements under which a project can be carried out in the HMRC Datalab

• Applications to use the Datalab are project specific

Next steps

Consolidate the success of the Datalab:

- Continue to improve the experience of Datalab users
- Promote the Datalab amongst the research community
- Encourage discussion between HMRC and the research community
- Ensure successful collaboration with the Tax Administration Research Centre (TARC)
- Continue to collaborate with other Data Services and international initiatives such as Data Without Boundaries
- Outline the role of the Datalab in the context of the Administrative Data Service and Administrative Data Research Centres
- Accreditation

HM Revenue & Customs

Thank you

Datalab team: hmrc.datalab@hmrc.gsi.gov.uk