Knowledge, Analysis, and Intelligence (KAI)

Nature and use of datasets held in HMRC's Datalab

Background

1.   The HMRC Datalab was launched in May 2011 as a secure environment where accredited researchers from independent institutions can access, free of charge, anonymised taxpayer and customs data, for research and analysis purposes that benefit HMRC's functions. Datalab research is not commissioned by HMRC.

2.   By making the taxpayer-level data available in a safe way, the Datalab supports knowledge sharing and policy making, facilitating new areas of research using improved methodologies whose findings should benefit society.

3.   Following a number of requests from external stakeholders about the nature of the information shared under this initiative, this note summarises the datasets held in the Datalab and the rigorous safeguards that are used to protect taxpayer confidentiality.

Data protection policies and standards

4.   The Datalab adheres to HMRC data protection policies and there are restrictions on working practices to safeguard confidentiality. In order to be granted access to the Datalab, research must serve an HMRC function under the Commissioners for Revenue and Customs Act 2005.

5.   Only approved researchers from trusted institutions can use this service. To become an 'approved academic researcher', institutions have to submit a proposal and complete a short training course.

6.   The rigorous safeguards in place include the data being anonymised to standards consistent with the Information Commissioner's Office Anonymisation Code of Practice before being placed in the Datalab. All direct identifiers such as name, address, telephone number and date of birth are removed and publicly available identifiers such as tax reference numbers are encrypted.

7.   Any output from analysis in the Datalab is subject to statistical disclosure control checks before release. In simple terms these aim to ensure that the outputs are sufficiently aggregated for no taxpayer's information to be directly or indirectly identifiable.

8.   Those accessing data in the Datalab are subject to the same confidentiality provisions as HMRC's own staff, including a criminal penalty to protect against unlawful disclosure of any information that could be linked to a named person.

Datasets available in the HMRC Datalab

9.   A list of datasets lodged in the HMRC Datalab is available on the internet:

https://www.gov.uk/hmrc-datalab-datasets-available

10.   HMRC also publishes a list of research proposals approved and the name of the institutions that have been granted access to the Datalab:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/367874/DatalabProposalsApproved_October2014.pdf

Further information

11.   Further contextual information on the Datalab is available from:

https://www.gov.uk/government/organisations/hm-revenue-customs/about/research

12.   Please contact the HMRC Datalab (hmrc.datalab@hmrc.gsi.gov.uk)