European Platform Undeclared Work

Data mining tools and methods, UK

Title of the policy or measure (in English)	Data Mining in HM Revenue and Customs (HMRC)
• Case study/good practice name	Data mining tools and methods to tackle the hidden economy in the UK
• Country	United Kingdom
• Date of publication	November 2017
• Sectors	All sectors
• Target groups	Direct target group: users of the tool, primarily HMRC staff Indirect target group: individuals and businesses at risk of undertaking a hidden economic activity
• Type of measure	Deterrence: improve detection
• Short sentence summarising the measure	HMRC employs a number of methods to detect hidden economic activities. These cover analytical tools (Connect), analytical methods (such as dynamic benchmarking and predictive analytics), and the acquisition and exploitation of Big Data (mostly focused on address matching). All these tools and methods combined seek to reduce the size of the hidden economy in the UK.
Background	
• Background context driving the implementation of the measure	The hidden economy is seen in the UK primarily as a tax issue. There has been considerable success in reducing the tax gap and HMRC and the Department for Work and Pensions (DWP) have also been successful in tackling fraud and error in benefit payments. The data mining tools developed by HMRC are part of these efforts. According to HMRC's own data, the total tax gap is made up of a number of behaviours ranging from the hidden economy, evasion, avoidance, errors (including criminal attacks), legal interpretation, non-payment and failure to take reasonable care. The data mining tool and methods developed by HMRC seek to improve tax compliance by reducing evasion, avoidance and the extent of activities that take place in the shadow economy. Overall, the HMRC approach to tackling the shadow economy seeks to promote a behavioural change for those in the hidden economy.
• When was the measure implemented? (including start date and end date/ongoing)	Data mining techniques have been used by HMRC for a number of years. Connect, one of its main analytical tools, was launched in 2010. HMRC works continuously in improving the various data methods it uses to tackle the hidden economy and increase tax compliance. As such this is an ongoing measure.
• Name of authorities/ organisations involved	Her Majesty Revenue and Customs (HMRC).
• Scope of the measure (pilot project, nationwide, regional)	This is a mainstream measure implemented across all administrative levels (national, regional, and local) where hidden economy practices may be taking place. It is also used to improve and increase tax compliance generally.

• Type of (policy) measure	It is a tool made up of a number of analytical methods, tools and data exploitation techniques to detect tax evasion.
• Key objectives of the measure	General objective: - To improve tax compliance generally and detect tax non-compliance across all sectors and activities of the UK economy. Specific objectives: - To reduce the size of the tax losses due to the hidden economy, -To reduce the size of the hidden economy itself, - To increase tax revenues.
Specific measure	
• Description of how the measure operates in practice	HMRC operates a comprehensive data analytical model within their compliance approach to tackle the hidden economy in the UK. This model is designed to process and analyse vast volumes of data. It is made up of a number of analytical tools (Connect), analytical methods (Dynamic Benchmarking and Predictive Analytics), and the exploitation of Big Data (through Address Matching Techniques). Connect is the primary tool HMRC uses to tackle Undeclared Work (UDW). Up to 4,000 users have access to the tool, which is able to perform 13 million searches. Connect is made up of 22 billion lines of data, and 500 million documents. Around 250 data analysts at HMRC are specialised in the Connect tool. The tool uses information from all HMRC data systems related to tax declarations for self-employed individuals, employees and employers, companies and business, property and land taxes, and indirect and consumption taxes and makes connections between the data to identify all data related to individuals and businesses. In this way HMRC is able to see a comprehensive picture of its taxpayers and the data, (HMRC and third party data) relating to them. Recently HMRC has been using the data within Connect to create maps of UDW, overlaying their data onto mapping software to provide a detailed visual map of UDW down to street and property level. They aim to use this approach to better target their compliance resource into risky locations. Dynamic Benchmarking is used to allow the data collected by HMRC in specific sectors of the UK economy to tell what the norm is with regards to tax revenues. This method serves to identify outliers to whom notification letters are sent. For instance, in certain sectors in the UK credit card transaction ratios can be used to identify outliers where the ratio between cash and credit card turnover is outside the norm. Predictive Analytics are used to identify the riskiest cases for intervention from VAT traders. The data taken from VAT population from: returns, debt information, trader characteristics and audit visit outcomes. The methods of analysis are mostly econometric multivariate models to predict those riskiest cases in need of targeted surveillance. Finally, the exploitation of Big Data is used by HMRC to trace down residential or commercial addresses at risk of undertaking undeclared economic activities. The process is largely automatised as it handles large amounts of data. However, human-in-the loop decisions are needed in order to solve potential pitfalls of the address matching process.

• Which groups are targeted by the measure?	The comprehensive nature of the data mining approach developed by HMRC to improve tax compliance means that it targets all economic agents (individuals and business) across all economic sectors. However, emphasis may be put on those economic agents deemed to be more at risk of performing all or part of their economic activity in the shadow. For instance, self-employed people relying on cash payments.
• What resources and other relevant organisational aspects are involved?	The measure requires a sustained commitment of resources to upgrade the IT (hardware and software) involved in its various analytical tools. Also, with regards to human resources, their high level of specialisation requires to commit resources to their specific training activities. There has been a significant increase in funding at HMRC to tackle the hidden economy which, through the Budget and additional internal deployments, will increase the total workforce (including managers and support) to over 1,000 staff. Cooperation with other government agencies is also developing to tackle the social consequences of illegal working and the hidden economy.
• What are the source(s) of funding?	National funding (public budget).
Evaluation and outcome	
• Has the measure achieved its objectives?	The objectives of the data mining tools and methods are: - To improve tax compliance generally and detect non-compliance across all sectors and activities of the UK economy, - To reduce the size of the tax losses due to the hidden economy, -To reduce the size of the hidden economy itself, - To increase tax revenues. Recent data published by HMRC shows a long-term reduction in the tax gap from 8.3% in 2005-06 to the latest figure of 6.5% in 2014-15.
• Assessment method (including indicators used to measure its impact), and the outputs and outcomes achieved	According to information published by HMRC in its annual report on Measuring Tax Gaps 2016, which breaks down the total tax gap by behaviour and distinguishes between the hidden economy, evasion, avoidance, errors (including criminal attacks), legal interpretation, non-payment and failure to take reasonable care, for the tax year 2014-15 the tax gap was 6.5% of liabilities (GBP 36 billion). The tax gap has reduced from 6.9% of liabilities in 2013-14.
• What are lessons learnt and the key conditions for success?	HMRC runs regular awareness-raising campaigns to encourage tax declaration, targeting high-risk sectors. These campaigns are very important as preventive actions alongside the activities carried out by HMRC within its Data Analytics department.
• Level of transferability (e.g. other countries/groups/sectors)	The transferability of the measure depends on the IT capacity of a country as well as the required level of IT skills of the human resources assigned to tackle the hidden economy by the responsible Tax Authority. Whilst many countries already have electronic databases for tax-related data, these are not always held on one central database, which may be due to legal reasons. Holding separate database systems is more costly and less efficient than using a combined system which is accessible to all users. However, some countries may require legislative changes before this can take place. With regards to ensuring the required level of skills of the workforce employed in the data analytics service of the tax authority, specific training provisions may need to be made in some countries.
Additional information	
• Contacts	*Mr Mike Molloy, Her Majesty Revenue and Customs, Assistant Director – Data Analytics.*

	Email: mike.molloy@hmrc.gsi.gov.uk
• Sources	HMRC strategy (Corporate Report, July 2017) https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/629941/HMRC-Strategy.pdf
• Metadata and key words for online search	United Kingdom, HMRC, Connect, hidden economy, tax avoidance, data mining.