

Closing the Tax Gap for HMRC using a **Data Quality Framework**

Butterfly Data helped tackle a £32billion tax gap by developing a scalable data quality framework, using SAS 9.4, DMS, and Viya.

Challenge

In the 2020-2021 tax year, HM Revenue & Customs faced a £32 billion tax gap - the shortfall between taxes owed and taxes collected. A large portion of this gap was tied to poor data quality, which hampered anomaly detection, compliance enforcement, and revenue forecasting. Several key barriers stood in the way: little visibility over where data issues existed, no standard metrics for assessing data quality, an Enterprise Data Model (EDM) that was inconsistently applied, reliance on a dated and under-documented tool (Data Management Studio, or DMS), and restricted access to live data, which delayed issue detection and remediation.

Solution

Butterfly Data designed and implemented a scalable framework to tackle these challenges, combining SAS 9.4, DMS, and SAS Viya 3.5. Data was accessed via SAS 9.4 and split into smaller, manageable segments to avoid overloading systems during processing. Custom, reusable jobs were created in DMS to enforce data quality rules across multiple datasets, drawing on previous work for the Metropolitan Police to accelerate delivery.

A new multi-layered Enterprise Data Model (conceptual, logical, and assumed physical) replaced the inconsistent legacy EDM, giving a clearer structure for data management and analysis. To make quality issues more visible and measurable, a scoring mechanism was introduced to allow both individual dataset errors and system-wide data quality to be tracked. Cleaned and scored data was then visualised in SAS Visual Analytics via the CAS engine, enabling high-speed, in-memory computation for responsive dashboards.

User stories shaped dashboard design, ensuring relevance to stakeholder needs, and an agile backlog with documentation was established to support continued improvement and facilitate handover.

 More on the next page



Ready to transform your data?

Book your free discovery call and find out how our bespoke data services and solutions could help you uncover untapped potential and maximise ROI.

hello@butterflydata.co.uk
+44 29 212 00140
butterflydata.co.uk

Get in touch 

Impact

Butterfly Data's work laid the foundations for a repeatable and scalable data quality framework within HMRC. A generic library of DMS jobs can now be reused across datasets, reducing duplication of effort. Processing large datasets in segmented batches has kept server load consistent and helped maintain stable performance. The dashboards make complex data quality insights accessible to non-technical stakeholders, enabling clearer, actionable discussions about where data issues matter most.

These tools and methods have enabled data-driven conversations amongst stakeholders about priorities for remediation. Despite constraints, such as limited tooling and data access, innovative workarounds like using dummy data initially, internal upskilling, and SAS partner training have helped HMRC better manage its data integrity. This contributes directly towards closing the tax gap through improved data quality.



Ready to transform your data?

Book your free discovery call and find out how our bespoke data services and solutions could help you uncover untapped potential and maximise ROI.

hello@butterflydata.co.uk
+44 29 212 00140
butterflydata.co.uk

Get in touch →

Our services



Data quality, governance, and privacy

Improve data quality, compliance and confidence in your data.



Data engineering and cloud adoption

Create scalable cloud platforms with robust data pipelines for multi-source integration.



Analytics and visualisation

Uncover actionable insights and make informed decisions.



Data science and AI

Deploy custom machine learning models and ethical AI to optimise, automate & enhance.