



A Data Integration and Governance Use Case: Inter-Agency Data Exchange

WHEN GOVERNMENT AGENCIES collect data on their citizens, there is understandable concern. There are more stories about data breaches and misuse than benevolent and helpful ones. That and a lack of citizen control and government accountability over big data is feeding mistrust and calls for limiting data collection. But what if those agencies could share citizen data in responsible, beneficial ways?

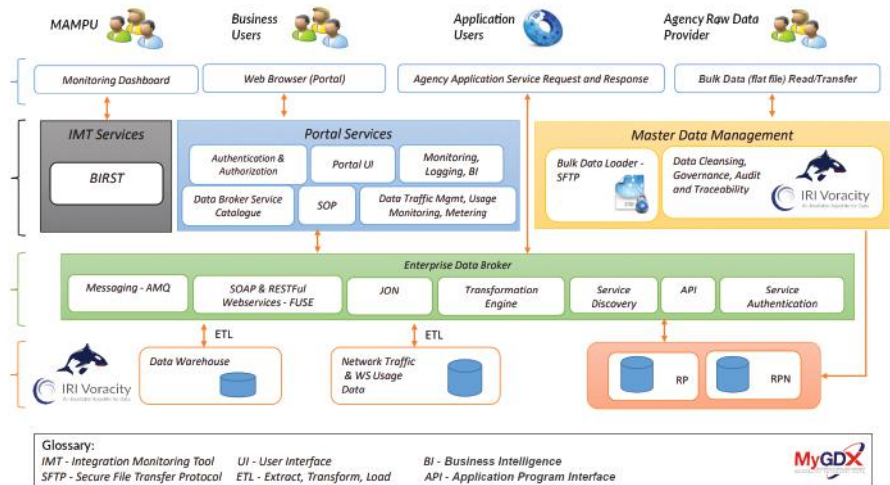
For example, many countries do not know how or why many of their students go on to college, choose certain careers, or remain unemployed. That's despite the fact that information on students with unique national IDs was collected in: primary and secondary schools (run by one or more agencies), colleges or training centers (reporting to other agencies), and military, healthcare, and penal institutions (still more agencies).

Correlating this siloed data could reveal which students from specific areas or backgrounds achieved specific outcomes. This in turn supports hypothesis testing, factor identification, and intervention strategies to improve those outcomes. And that is only one example.

Enter an integrated master data management system for government agencies to exchange data with each other for public good ... one that leverages best-of-breed data storage, integration, governance, and tracking technologies to ensure that heterogeneous agency data is unified, de-identified, and shared properly.

MYGDX & VORACITY

The Malaysian Government Data Exchange is a state-of-the-art web portal supporting the uploading, integration, downloading, and access-logging of de-identified data in multiple government agency silos. MyGDX holds the promise of producing more "unified views of the citizen" that not only improve early



Architectural diagram of governed ETL and MDM in MyGDX via IRI Voracity

intervention strategies, but also could spot disease clusters, prevent crime, and better target public and private services.

MyGDX is front-ended in a modern web portal that delivers available but de-identified data while tracking all requests. It will also support monetization of the aggregate data for commercial use.

To meet the system's quality, privacy, and analytic goals, that data must be rapidly integrated and reliably governed at once. And to do that, MyGDX uses the combined data integration (ETL), unification, cleansing, and masking functionality of the IRI Voracity data management platform.

Voracity jobs are powered by the long-proven IRI CoSort data acquisition, transformation, cleansing, and masking engine. Those jobs are built and managed in IRI Workbench, a rich and familiar IDE supporting job design and workflow management in multiple modes, built on Eclipse.

COMBINING DATA DISCOVERY, INTEGRATION, MIGRATION, GOVERNANCE, AND ANALYTICS

Data integration and governance have been traditionally separate disciplines

performed in separate silos and products, and managed by people with different agendas and skills. Think about all the separate and costly data profiling, ETL, data cleansing, data masking, master data management, and IAM/auditing technologies that your company may be entertaining.

Voracity is the rare platform product that seamlessly combines these and other key data lifecycle management activities in the same pane of glass and I/O pass.

Inside IRI Workbench are multiple, fit-for-purpose wizards that build and/or execute portable jobs to search, classify, and apply consistent mapping, masking, and test data generation rules. Also included and seamlessly interlinked: data profiling and classification, single-pass ETL, data quality, PII masking, metadata management, analytic options, and interchangeable Hadoop runtime engines to handle on-premise and cloud data sources, big and small. ■