Paul Carey

Data Management at HSBC

Simplify Complicated Data Management Schemes at HSBC

SEPTEMBER 3, 2020

Have you experienced an organisation with overwhelming amounts of data that never served a purpose? Sometimes it can be a form of vanity, with managers happily touting collected information without purpose. In other cases, a truly visionary leader can know the data will eventually be necessary. They understand that one day this information will be useful, and when that day arrives, they'll need it.

In any case, one of the more recent revelations is the benefit of the complete decentralisation of data. Modern leaders have long known that they needed data. What we are learning now is just how powerful full data access is in the hands of the entire organisation.

Data quality, for example, is easier to achieve when the entire organisation knows its role in maintaining data integrity. Making data-driven decisions is also simpler when the insights from the analysis can be distilled into actionable lists for frontline employees. Companies where everyone contributes to business intelligence are better able to steer towards success.

HSBC's Approach to Data Management

HSBC was founded in 1865 to facilitate international business between Europe and Asia. Initially opened in Hong Kong, our bank now serves more than 40 million organisations and individuals around the world. We realise that thriving through the next century will require a strong commitment to analytics.

My educational background is in information technology and systems thinking, but I spent most of my career involved in the financial sector. As a data management specialist, I am well acquainted with the struggle to wrangle data complexity and scale into a usable format. It is not enough for data experts like myself to print out reports for organisational leaders. Truly achieving the promises of data means creating a data management path that makes the process simpler.

To unlock the power of data, strive for simplicity.

At the staff level, that means helping our entire organisation to take ownership of data management. Every single person with data access needs to have an easy way to report data quality problems. Because HSBC is such a large company with a massive commitment to data collection and analysis, we had to boil data management down to five key concepts. That helped us better manage our approach to organisation-wide data policies and procedures.

The first key concept is: **what data do we care about?** Obviously, this is a fundamental step in any strategic data analysis. Unless you just like collecting data like stamps, you need to ensure you're collecting information that can directly help your decision-making. Then we need to know: **where does it reside?** This information is especially important given the sheer size of the HSBC operation.

The third concept we need to address is **why we leverage the data from "that" place it's collected**. In other words, we need to make sure the data we collect is trusted for a particular business process. The fourth concept relates to the **impact/opportunity collecting this data may have on the organisation**. That can help us to establish some priorities that can inform resource focus. Finally, we need to **ensure the data is fit for the intended purpose.**

Using these five concepts to guide our data management decisions helps us keep our processes efficient and valuable. We developed lineage diagrams that clearly linked each data point to a business purpose and location. We could also clearly see the impact that data may have on the relevant business practice and use all this information to develop data quality mechanisms to keep the data fit. Our dashboards are created with these five concepts in mind.

Following Data Through Its Lifecycle

Once we have links between the data and the business uses, we can better chart the life of data within HSBC. Understanding the lifecycle stages of our data enables us to better create a series of dashboards. These dashboards can capture each stage of the data lifecycle and make sure information can be readily conveyed throughout each level of the organisation. That's where we rely on data analytics platforms like Qlik.

Qlik makes dashboarding incredibly intuitive. Data is less than useless if it cannot be properly shared and contextually analysed.

The first step to making the data more coherent and easier to navigate is a data dictionary. Because we are such a geographically large organisation and have staff with a variety of backgrounds, we need common definitions. Having these definitions available makes it possible for staff throughout the HSBC to speak in common terms. It also makes it simpler for separate teams to work together on common goals.

You can't leverage data properly without common definitions.

Within our organisation, we also use our data dictionary to categorise the data and convey the business use in context telling us where the data is currently being held and implemented. Finally, the dictionary also includes information regarding data quality requirements. Including data quality information as part of the initial definition makes it easier to help all staff share responsibility for the integrity of the information.

The next logical step once the data has been defined is providing an inventory of applications. This inventory includes basic information about the applications, such as purpose and ownership. The ownership piece is included so that we can know whom to contact in case there are problems. Most

importantly, the application inventory explains what data the app is producing and where it is being distributed. That allows us to track data across the world and through different parts of the organisation.

In the business process repository, we address *why* the organisation collects this particular data. That allows us to answer business process questions and fully understand how the data is used. In a perfect data-driven organisation, every single business process would be listed in this repository along with relevant data.

The final two stages involve data quality. Now that we understand the why, where, and how stages, we need to address the quality. This is how we make sure the data can serve the purpose for which it is collected. With customer addresses, for example, the data is only useful if it is complete, accurate, and up to date. We also don't want a bank in one country to send a note about new regulations to customers in a different part of the world because the location is mislabeled. The final stage defines various automated quality monitoring tools in place.

All of these stages culminate in a utopian data management vision. In a perfect world, we would always have a real-time view of our data whenever it is needed. We also need to provide people with the ability to access that data everywhere, and not just at particular machines. All this has to be sustainable and ongoing because we will need data five years from now for the question we don't know how to ask today. That is why we call it data management, it's an ongoing process.

Pulling It All Together

The Data Management Summary dashboard we've created with Qlik provides an example of how we pull it all together. The summary of all relevant data is linked to the five concepts mentioned earlier.

Beginning in the centre of the dashboard page is the list of all the data elements, business terms, and data categories. These explain exactly what data is being collected. On the left side of the page are the Scope Dimensions and other filters. These allow you to limit the data by where it is relevant. Users can filter data by geographic locations, business processes, and business functions.

Across the top of the dashboard are descriptions of the scale. We can view measures like the number of data points, business processes, data fields, and business terms represented on the page. Of course, higher-level executives will have larger numbers to analyse. As you drill down using the various scope menus, you can get specific information relevant to a particular office, country, function, or product. The more you drill down into the data, the smaller number of represented data points, business processes, and so on.

The page also has several features related to data quality. At a glance, the user can monitor data quality, data governance, and data architecture numbers on the top. Each of the data rows is also colour-coded so that you can easily match countries, business processes, and other data categories. Some data points may be listed in red to draw attention to data quality issues.

At the end of the day, we appreciate how easy it is to create a page that allows us to represent so much information with clarity. It is not enough that your data specialists are able to easily read and manage your data. Your entire organisation must take responsibility for data management. But they need an easy format to make that possible without causing too much frustration.

With simplicity and business value focus, comes adoption.