


Data Quality Improvement

START SMALL, THINK BIG

A GovTech eBook by

DATACTICS™



Like the weather in Britain, poor data quality is often seen as an intractable problem. Why should that be the case?

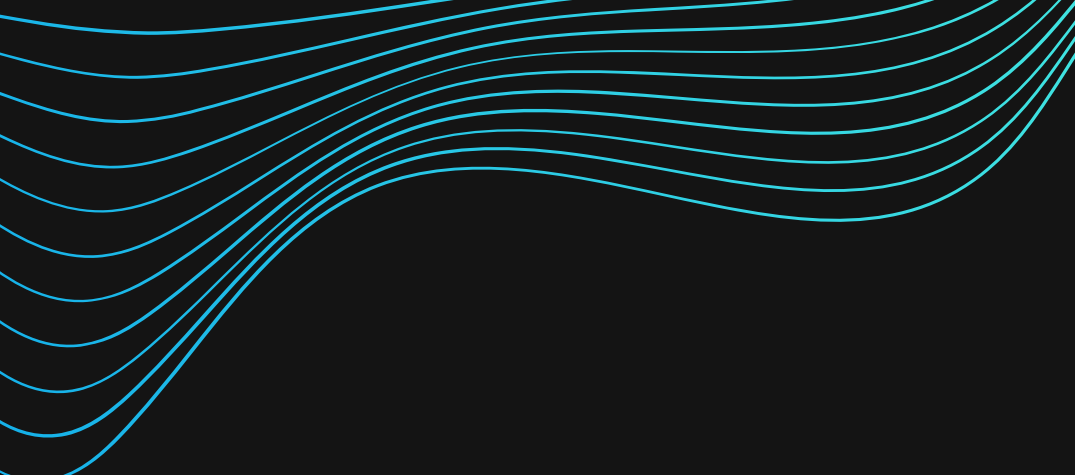
At Datactics our experience, in a range of complex data challenges for highly regulated industries such as banking and government, is that with the correct combination of tooling and an intelligent approach to governance, data quality improvement can be achieved in an fast, iterative and cost efficient way.



Data-driven policing: a new imperative

Increasingly forces are being challenged to improve data analysis and insight. They require tools that assist with data preparation and data wrangling to provide clean, de-duplicated and well-structured information to multiple critical business functions. These include upstream analytics software, such as PowerBI Dashboards for management reporting, data science tools such as MatLab or R supporting predictions on crime trend demographics, and internal reporting to better interrogate the reserves of crime data they hold.

CHAPTER 1

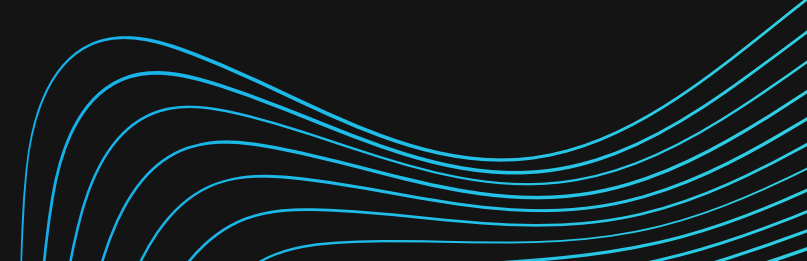


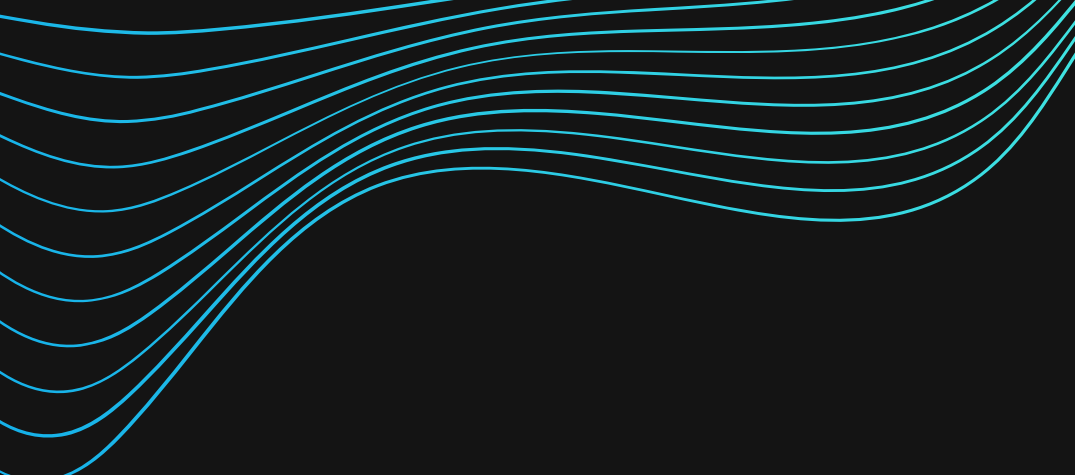
In their forward to the NPCC's "Digital, Data and Technology Strategy 2020-2030" the heads of National Police Chiefs Council, IMORCC and APCC stress that information is the lifeblood of police forces and highlight the challenge of dealing with the data that will support intelligence-led policing in the next decade.

Police forces across the UK will face many demands in relation to the volumes and types of information related to increasingly sophisticated crimes. Simultaneously police data stewards will meet increasing demands for transparency and accountability regarding their decision-making based on this data.

Constabularies currently face many challenges arising from the very large volumes of potentially messy information relating to people, objects, locations, and events ("POLE" data). Poor data quality, limited data standardization, validation failures and data duplication all get in the way of effective incident analysis and reporting.


Police officers spend too long retrieving data from silos and reconciling it before it can become actionable intelligence. Poor data quality can also create downstream problems for citizens and regulators as they conclude that an organization's data sometimes can't be trusted.





In the 2020 RUSI Annual Security lecture, Dame Cressida Dick asked her audience to take away the following message – “Policing will remain an essentially human service, supported by better information and tools”. The key message of these recent reports is that police forces should aim to unlock the power of data and put actionable information in the hands of officers and staff when it is needed. The key to the success of future policing will be to provide frontline staff with up-to-date and accurate insights from many disparate sources of data.

In addition, Dame Cressida referenced RUSI’s “Data Analytics and Algorithms – towards a new policy framework”. The report makes strong recommendations to government, regulators, police forces and software developers especially about how to meet society’s expectation about how we handle citizen information in the challenging new world of AI/machine learning.

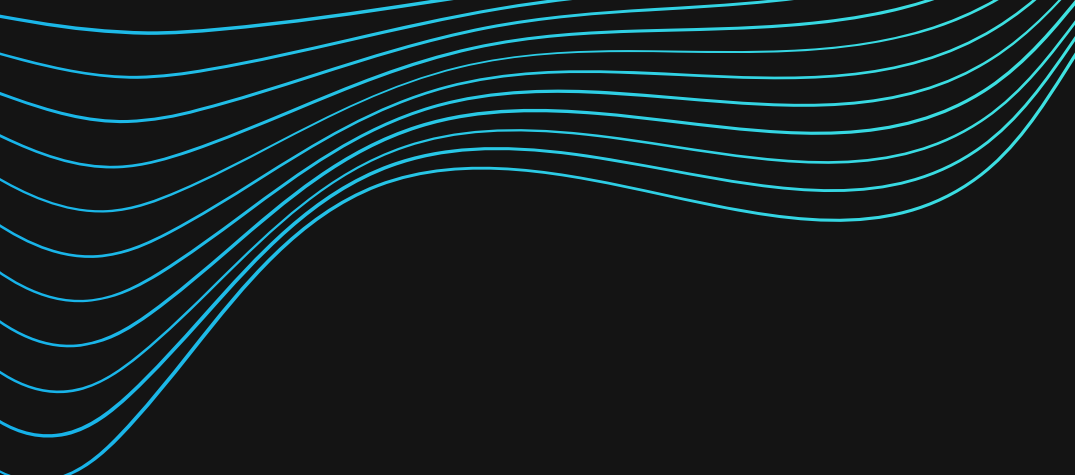




People, processes, and platform: the holy trinity of data quality

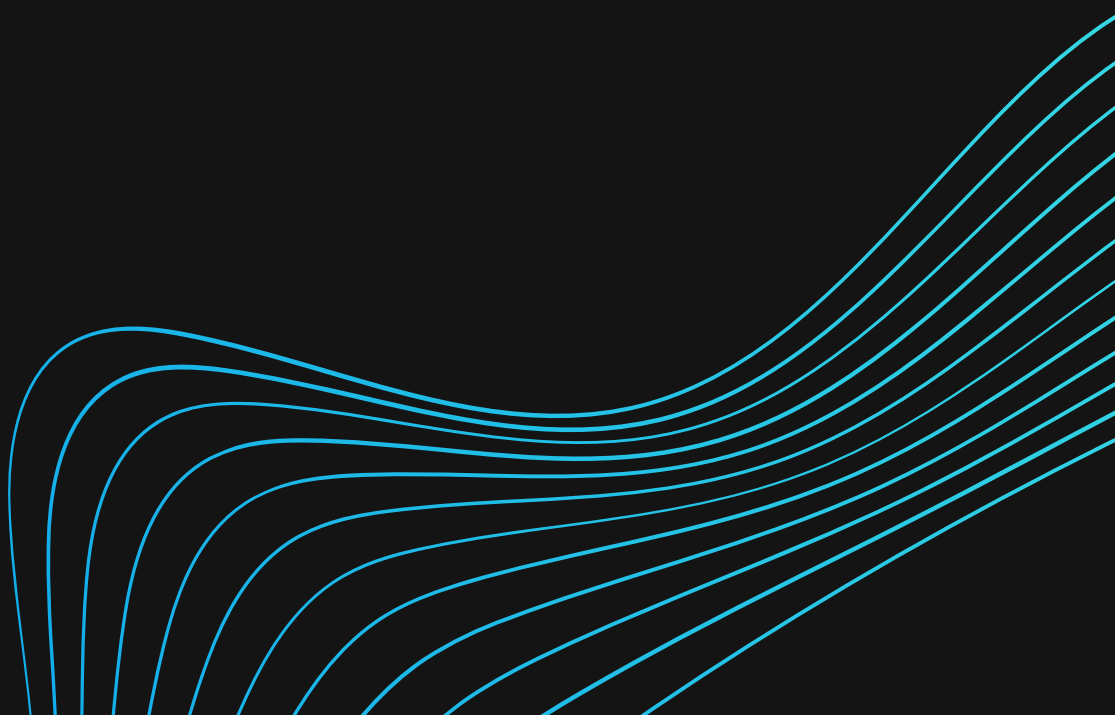
What part does improving data quality play in all of this? In this article, we discuss how improvements in data quality involving people, platforms and processes can assist in generating better data to feed the major information processes used in policing.

We suggest an agile approach that looks to augment existing reference data systems with compact and affordable tooling to measure and improve data quality as well as matching. We also discuss the essential data quality personnel roles required to make such tools effective and propose a range of practical use cases that might be attempted by any force.



In our experience, successful data leaders should measure quality, create and staff key roles, pay attention to tooling costs and explore the art of the possible via a well-defined proof of concept (POC) where the client can easily demonstrate the benefits to their business. Ideally, this POC would start small and then be used to evaluate the costs and benefits of a broader data quality programme, while simultaneously building stakeholder confidence.

A successful POC should result in realistic timescales for a broader project and provide measurable indicators for the value of data quality improvements.



From our experience in delivering such programmes, we see the following potential use cases where measuring and improving data quality can be used to improve operational excellence in policing.

Police use cases

DATACTICS™



Single Citizen View

Golden record creation from multiple systems or silos in preparation for analytics and migration.



Self-service DQ

Continuous measurement, monitoring and improvement of data quality for compliance with regulations / standards.



Match Engine

Map and match individuals / entities / assets at scale to streamline on-boarding processes.



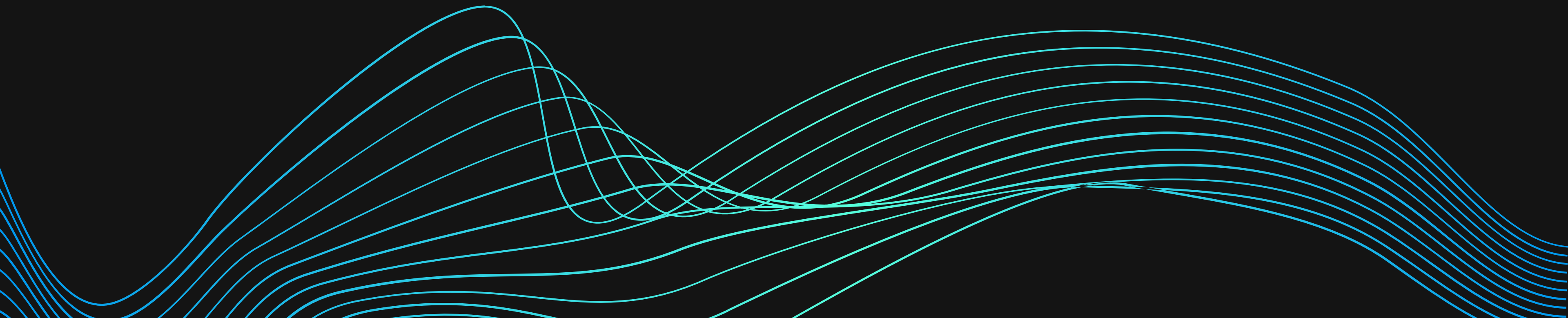
Data Migration

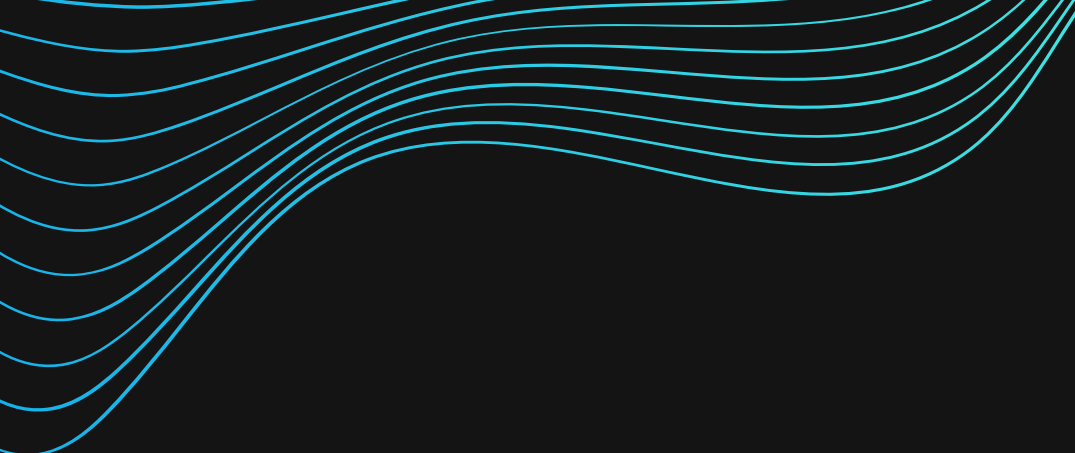
ETL + Data Quality. Profiling, validation, re-engineering, matching & de-duplication to ensure data integrity.



Self-Service Data Quality for DQ Metrics & Improvement

USE CASE 1





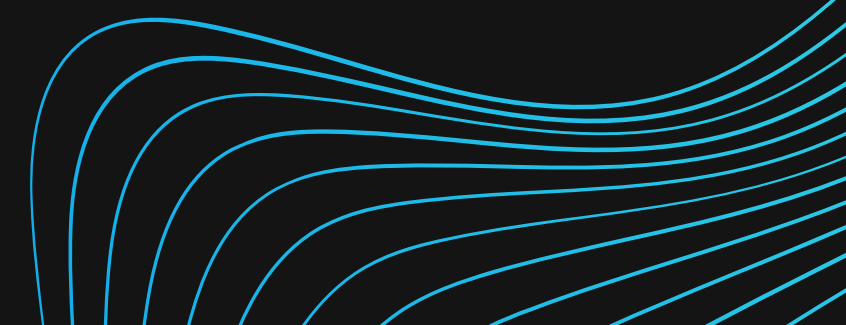
It is a fundamental tenet of engineering that “you can’t manage what you can’t measure”. Self-Service Data Quality allows for the continuous measurement and monitoring of live police data assets according to recent POLE & MOPI standards.

Data sources could be many and various including RMS and POLE data stores and holding information related to crimes, custodies, intelligence, case preparation and regulatory reporting.

Business users are typically interested in data quality metrics such as consistency, validity and completeness

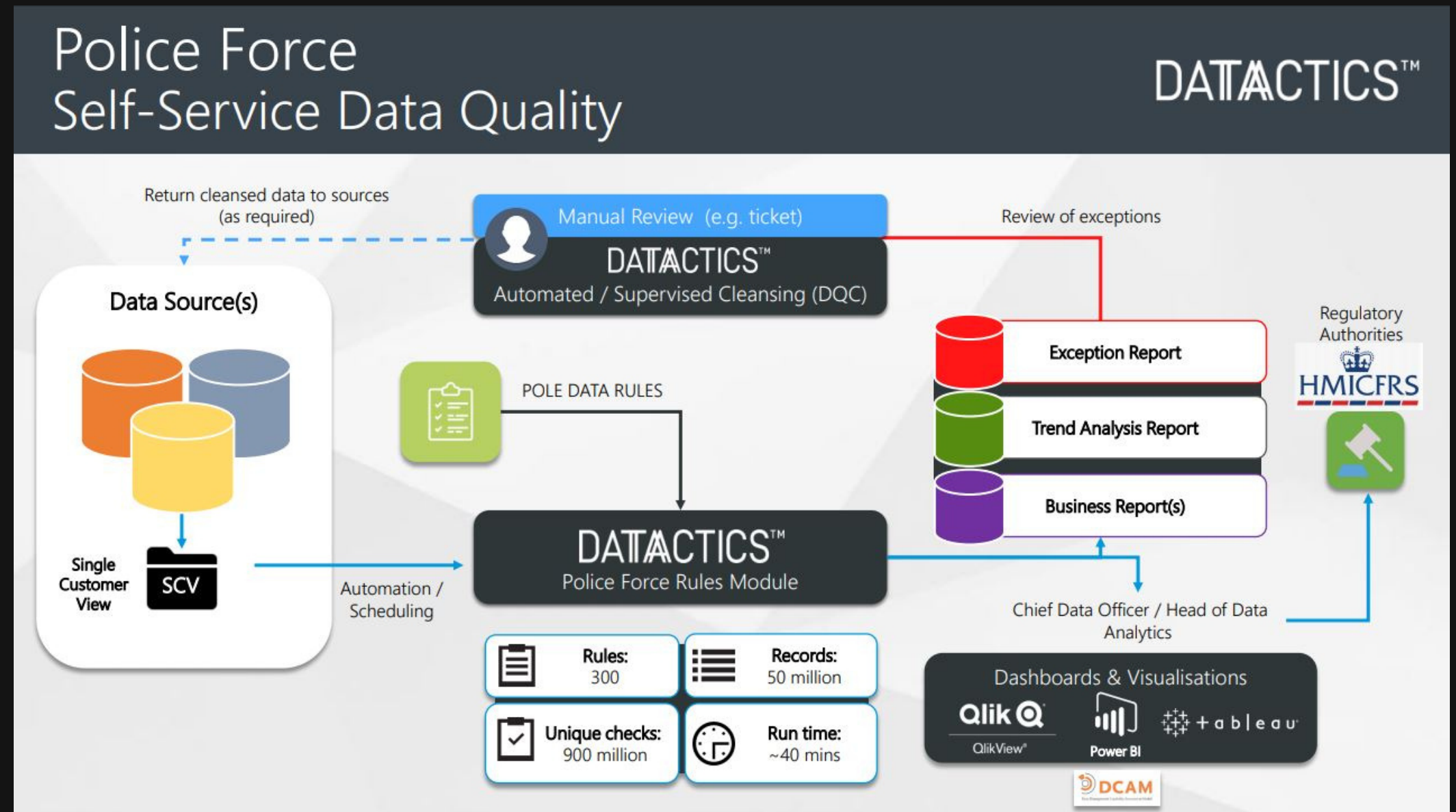
Rules relating to these data sets could start with simple accuracy checks (e.g. does a postcode have a correct format, or is a suspect’s age between certain limits) and then build to more complex logic involving gazetteer data where an address can be checked against a postcode or GIS reference.

The system has a range of connectivity options for third party sources of validation information relating to name and address, including open data from Post Office PAF and Gazetteers for GIS.



Metrics showing the improvement or degradation of data quality over time are available for review by data stewards via an easy-to-use dashboard and allow drill-down to failing records for correction if required.

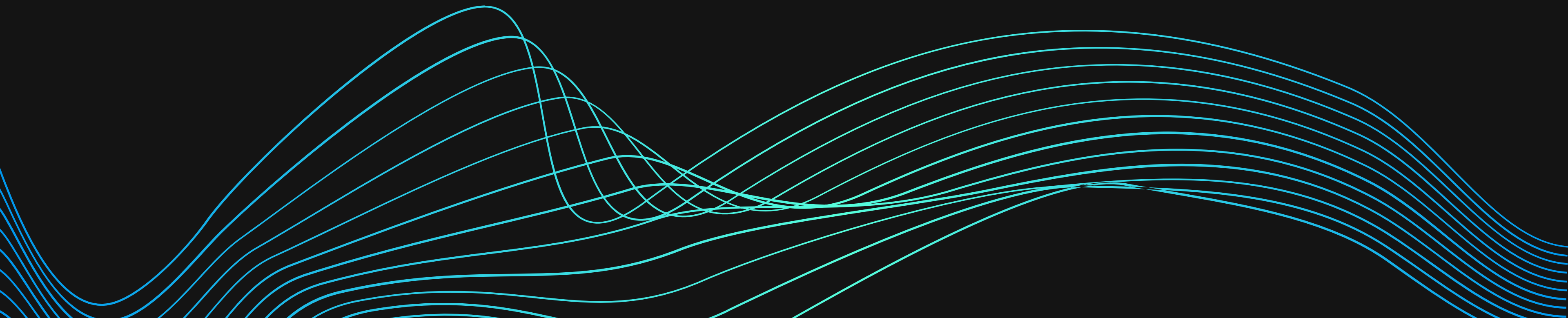
The Datactics system is designed to be operated by subject matter experts (SMEs) who can fine-tune out-of-the-box rules without needing to be a programmer or coder.

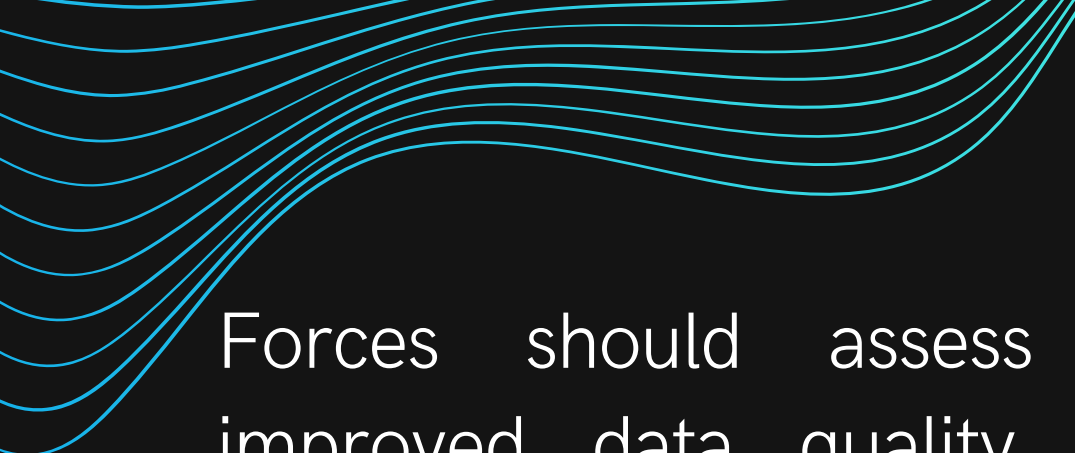


These SMEs can be force data analysts or external consultants, depending on whatever approach the force requires. Ultimately the purpose of Self-Service Data Quality is to allow end-users to monitor and maintain high quality data that is accurate, consistent, and fit for purpose.

Creating a Single Citizen View


USE CASE 2



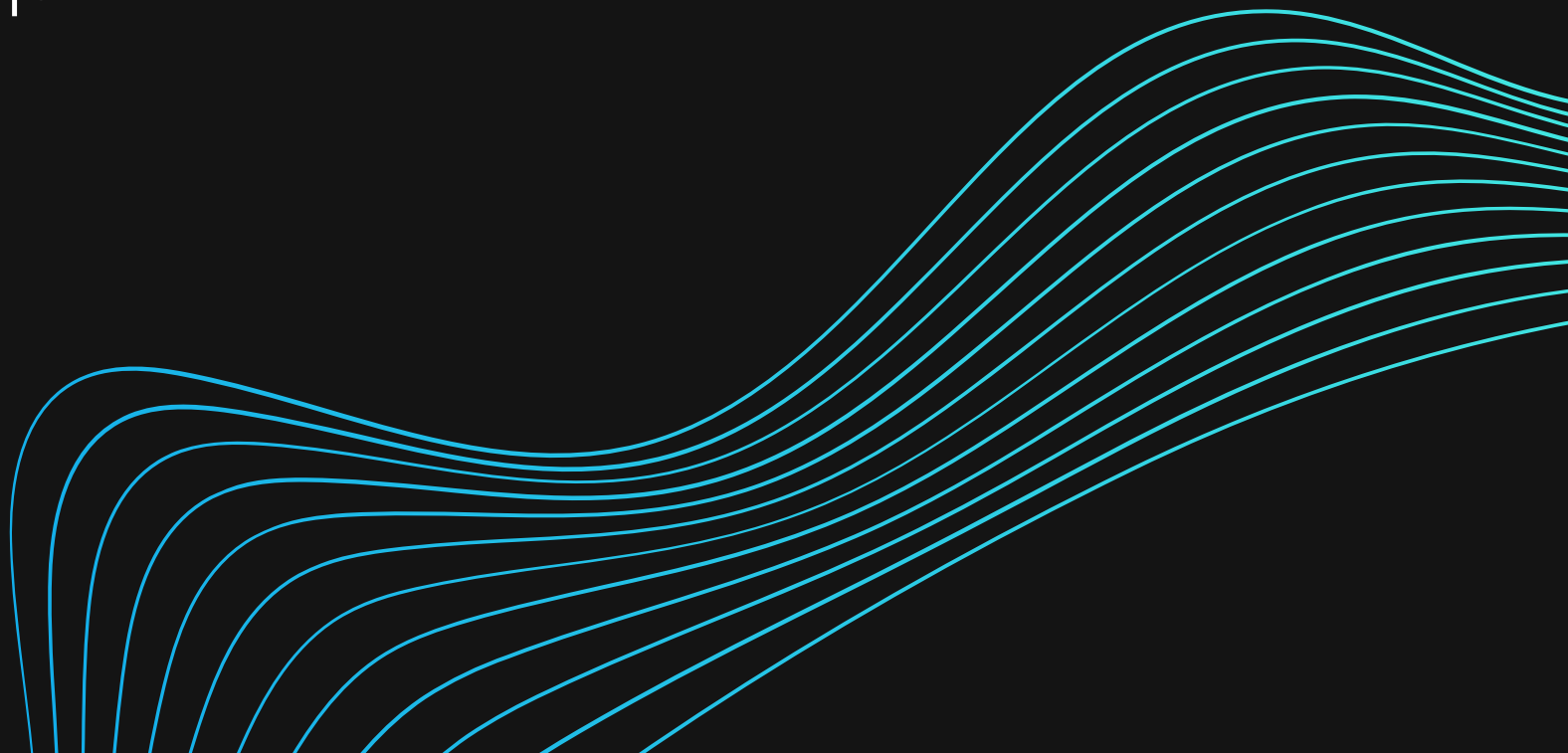


Forces should assess the benefits of improved data quality in regard to the duplication of police nominal data to create a “single citizen view”. Much like in banking and finance, where being able to see one consolidated view of a customer makes for better intelligence, marketing and regulatory reporting, this will vastly improve the activities that police forces can undertake based on better quality data.

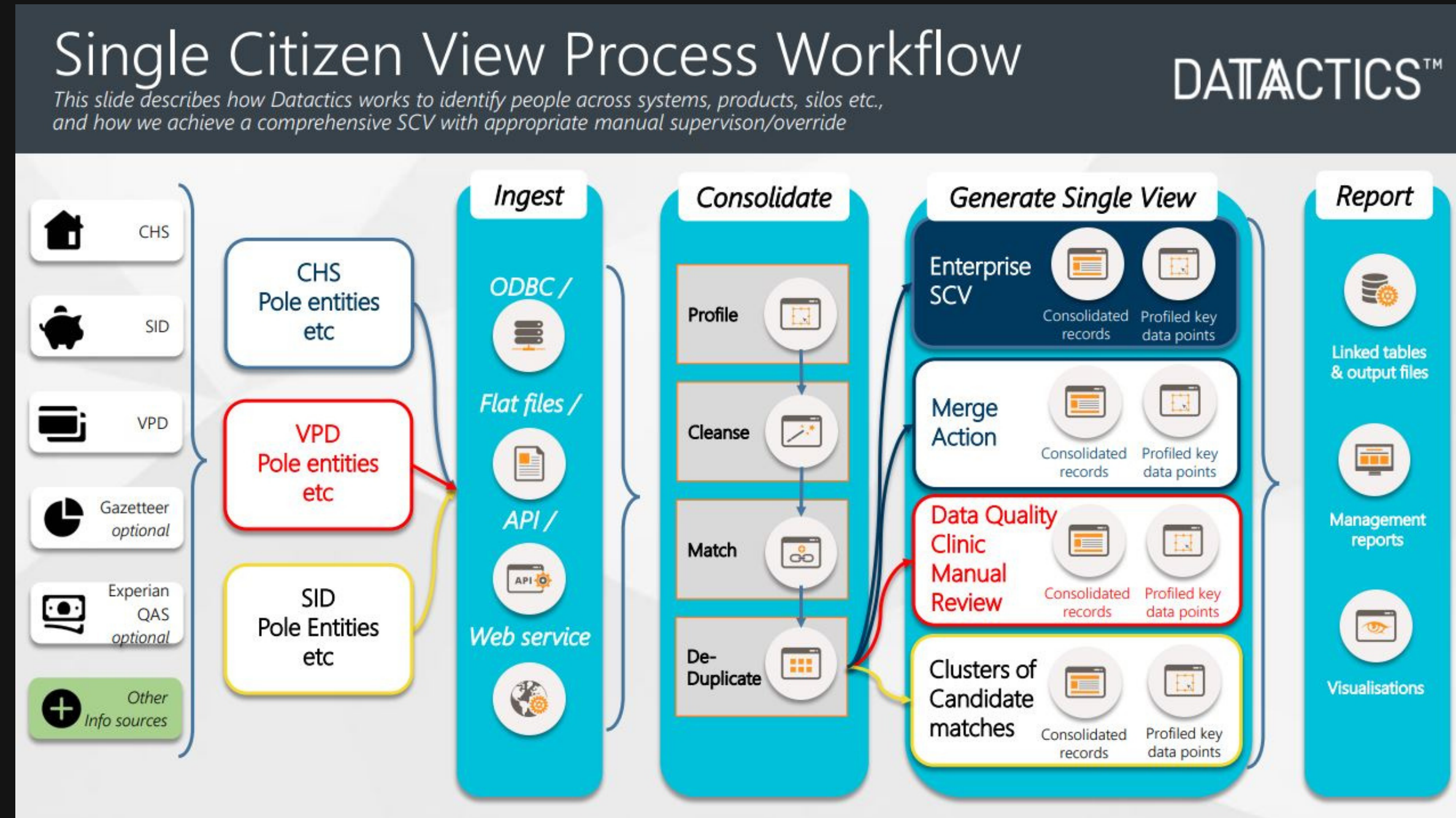
It would enable forces to meet regulatory reporting for GDPR, or to provide clean data into software used for predictive analytics which might otherwise be made less accurate by noisy data, for example containing duplicate nominals.



Creation of a single citizen view has several prerequisites. Firstly, it involves powerful matching on large data sets and de-duplication logic that allows for highly configurable fuzzy matching on name and address information. Secondly, the output needs to be one single golden record or a cluster of candidate records scored for how closely they match (known as a likely match rate). This process should not discard any previous metadata that can be used to understand the history of the person in question.



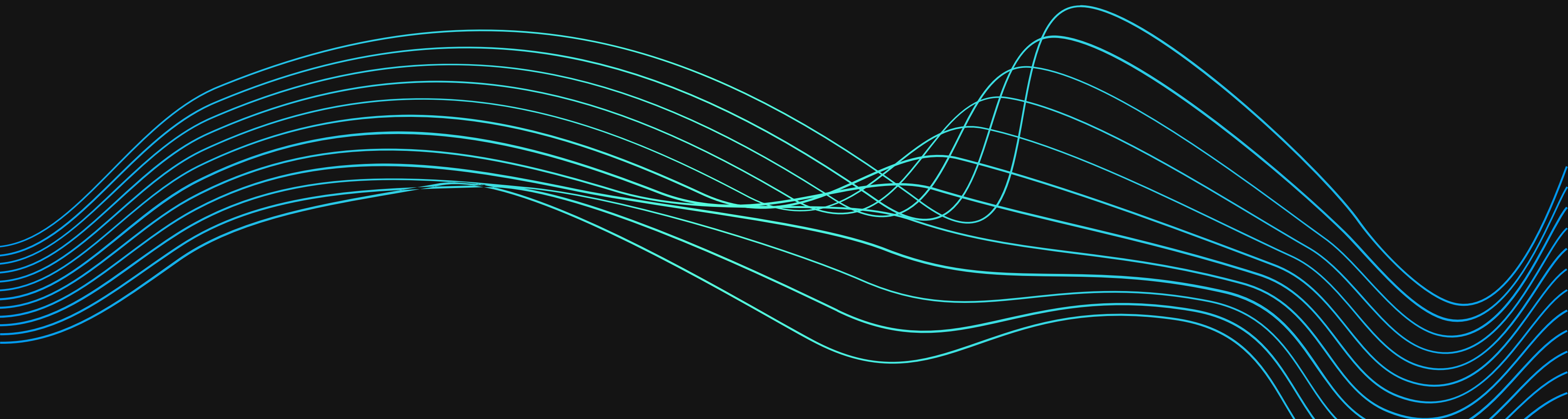
Additionally, matching should deal with issues regarding data falsification where a suspect may provide inaccurate information to avoid detection.



Challenges in this area include suspects deliberately obfuscating names via abbreviations and transpositions, as well as having to deal with multiple languages such as Chinese, Japanese, or Cyrillic scripts.

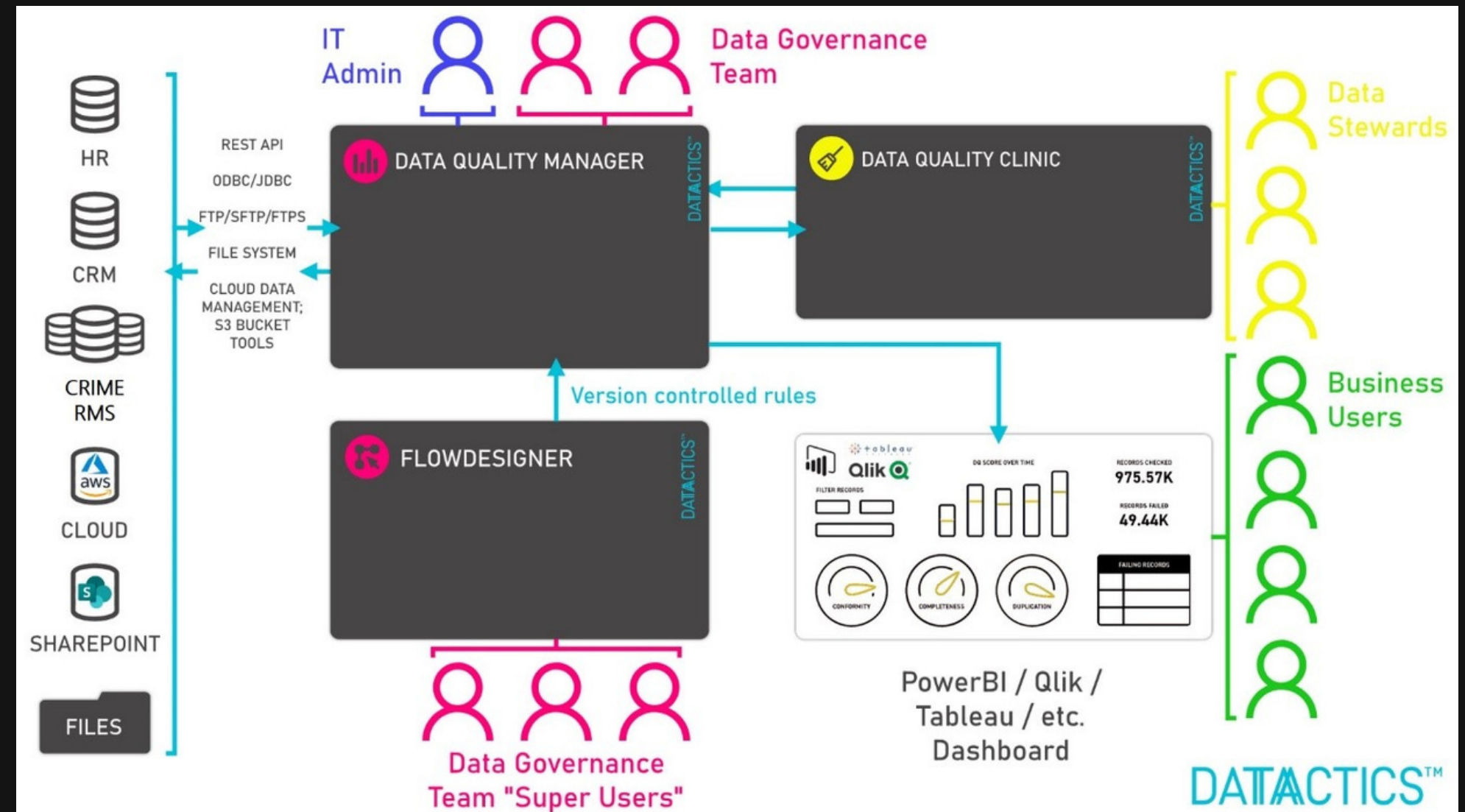
Data Migration using the Dataactics Platform

USE CASE 3



Many forces are modernising core technology. They are involved in the migration of data from legacy systems to next-generation solutions often for crime record management while aligning themselves with a national vision for police data and technology.

Historically these legacy systems contain large volumes of duplicated or incorrect information relating to crimes, custodies, cases, VDRS and HORT.



Ideally, data migration involves not just data transfer but allows for "Extract, Transform, Load (ETL) plus data improvement" as part of the operation. This will result in the new target system being populated with scrubbed and de-duplicated data and operating at a higher level of accuracy due to improved data quality.

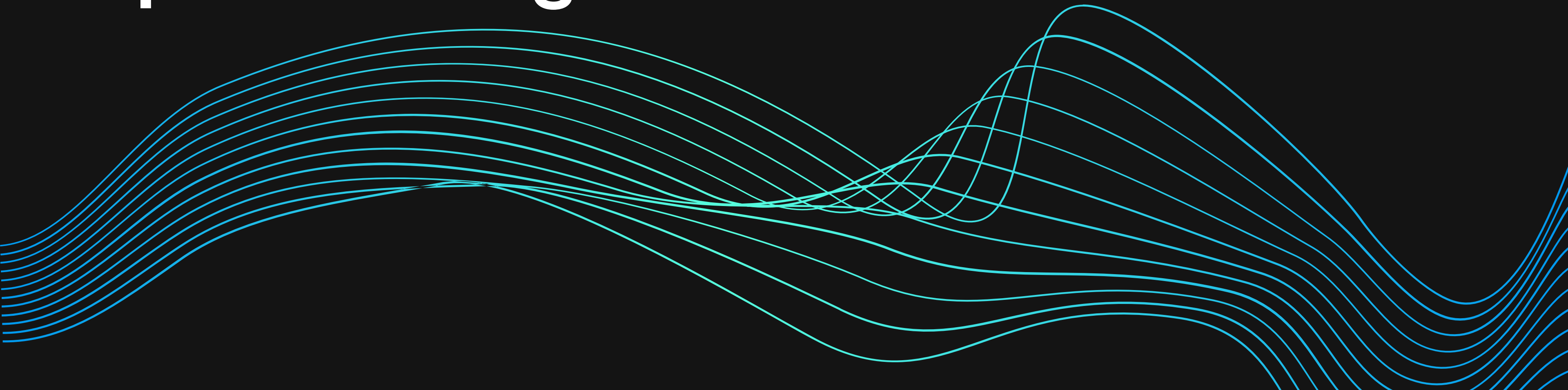
Start small, but make a start!

In the conclusion of National Policing Digital Strategy, the authors outline a “big picture” where data-driven insight has the possibility of being a “force multiplier” in terms of the contributions from multiple data sets, and the pace, predictability and precision of police work. They go on to emphasise the potential of big data and artificial intelligence as game-changers for policing in the next decade.

All these technologies require high-quality data that is clean, complete, and de-duplicated, measured to a high standard, and monitored over time.

Taking the first step with Datactics builds on our work already completed in the sector and within other highly regulated domains such as financial services and insurance.

Demanding better data to power excellence in policing is the perfect place to begin.



Thinking big?

Take the first step and get in touch:

Stuart.Harvey@Dataactics.com

Kieran.Seaward@Dataactics.com