

NEW FRONTIERS AND CHALLENGES
OF DATA GOVERNANCE

EXPERT OPINION



By Cyril Cohen-Solal | Director of Keyrus Innovation Factory, Keyrus



and Xavier Dehan | Director of Big Data & Analytics Business Development, Keyrus

Cloud and SaaS applications are being adopted more and more all the time, embedded analytics is becoming widespread, Big Data approaches are increasingly being democratized... In a context in which data flows are growing more intense, and the activities of business functions are increasingly reliant upon them, the quality and consistency of the data disseminated are becoming essential. So, who does it really fall upon to take care of this? There are new, shared tools which, by restructuring the roles of the business functions and IT, make it possible to achieve overall and sustainable data governance¹.

According to recent studies², companies estimate that on average one third of their data are inaccurate. 91% of them think that these incorrect data lead to inappropriate decisions adversely affecting their economic performance. Finally, 79% of organizations throughout the world think that by 2020 the majority of commercial decisions will be taken on the basis of customer data. Whilst there is nothing new about this assessment, the question of responsibility for data quality remains entirely relevant. Moreover, this question is becoming all the more complex since the data that have to be managed are increasingly to be found outside the company, beyond the direct control of IT teams. The challenge here is a critical one. At the end of the day, it is all about building up, enriching, preserving, and sharing a high quality data asset base: an essential asset for valorizing the company, ensuring its development, and guaranteeing its long-term future. As early as the 1990s, many organizations started adopting verification and validation strategies so as to ensure that their applications were "well developed", that they worked, and that the companies had made "the right product" corresponding to users' requirements. Business functions and IT worked together to organize these activities, notably by deploying methods, ways of organizing, and steering tools. Like the expression – in

the form of a statement relating to software qualification – "tests that were too few and too late!", here, we find the principle of "Garbage In – Garbage Out"³. Today, it is wise to adopt the same strategic approach, not to ensure the quality of the applications developed (of their code and specifications) any more, but rather to control the quality of the data. These two risk factors (the introduction of software anomalies and/or of incorrect data) contribute to the overall poor quality of the enterprise's IS. They also give rise to additional costs (those of a cycle to correct defects) and drops in performance (resulting from a cycle of inappropriate analyses and decisions).

THREE DEVELOPMENTS THAT ARE CHANGING THE SCOPE OF DATA GOVERNANCE

- **The success of business applications in SaaS mode.** The growing range offered by editors and the reductions in costs associated with SaaS solutions have got the better of companies' initial reluctance with regard to the Cloud. For example, in an area like CRM, which is sensitive because it involves customer data, since 2014 new deployments have mostly been in SaaS mode and generally undertaken at the initiative of the commercial and/or marketing departments. The fact that the data in SaaS applications are managed

¹ Definitions of Data Governance:

• by the Data Governance Institute (DGI): "Data Governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods." http://www.datagovernance.com/adg_data_governance_definition/

• by the Cigref: "Data Management consists of implementing all the set-ups relating to the information used within our organizations so as to optimize its use... data governance is the part of this which describes responsibilities, sets rules, and checks that they are applied. This governance approach is led by a dedicated body that steers it, with rules, guides, repositories, indicators, charters, a policy (relating to the management of personal data, classification, storage, retention, patents, intellectual property,...)"

"How can the enterprise's data be managed to create value?" Cigref Report – The Business Challenges of data <http://www.cigref.fr/wp/wp-content/uploads/2014/10/CIGREF-Enjeux-business-donnees-2014.pdf>

² "New Experian Data Quality research..." <http://www.experian.com/blogs/news/2015/01/29/data-quality-research-study/>

³ "Data Quality Benchmark report 2015" "Data Quality and Management – Market Trends in 2016"

³ GIGO: "incorrect data going in, incorrect results coming out"

according to rules not controlled by the company inevitably causes IT departments problems in terms of integration and consistency with internal system data.

- **The development of embedded analytics and in-Memory.** Most business applications now incorporate advanced analytics, visualization, and reporting tools that take precedence over former, centralized Business Intelligence systems. The latter also face competition from in-Memory analysis tools, which are less complex to implement and use. Whilst the adoption of these systems has the advantage of making business teams much more autonomous, it also results in a proliferation of repositories which nobody is able to keep consistent.

- **The boom in Big Data.** After years of paying lip service to Big Data, companies have finally begun to utilise not just their growing volumes of operational data, but also the gigantic sources of multi-structured data that are the social networks and, more broadly, the social web. Mastering the rules for collecting and transforming these data proves crucial to the relevance of the algorithmic processes (360° views, predictive analyses, real time requirement,...) that are then applied to them and it dictates the value of the results obtained. As of today, companies that have extended their data quality policy to cover these new flows and processes are few and far between.

The unified governance of data is to Chief Data Officers what the Theory of Everything is to physicists, such is the extent to which these two Data Worlds sometimes seem dissociated!

The main causes of this dissociation lie in the very essence of the Big Data revolution, with its highly experimental dimension which specifically affects data governance.

Indeed, Data Scientists work in a "free space". This is to allow them to constantly test their algorithms (analytical models and management rules, architectures, data ecosystems) by incorporating new sources of data (internal, external, semi-structured,) on the fly. Naturally, this activity is performed with limited traceability. Data Scientists are more focused on the extra performance points gained on their business or technological use cases (reduction of Churn, improvement in the prediction rate, real-time performance...), than on falling into line with the canons of data quality (accuracy, integrity, uniqueness, conformity, and completeness), data management best practices, or the pressing requirement for industrialization. Enterprises provide a partial response to this central problem by applying strategies of minimal data governance, to "contain" the risks with data quality and management, and by organizing infrastructures comprising dedicated, compartmentalized spaces for innovating, experimenting, firming up on PoCs, and undertaking pre-industrialization (Sandbox Lab...).

So, enterprises must meet the major challenge for the coming years of organizing the unified governance of their data. This covers reference data, which have to be unique, reliable, valid, and complete (customer name, address, products, assets...), decision-making data produced from transactional data (revenues and margins by BU...), and also all the massive data coming from Big Data, Cloud, and Digital infrastructures, which intrinsically comprise a degree of uncertainty. This is what enterprises have to do to be able to generate returns from their data capital. Today, that capital is a strategic asset for companies in terms of creating a competitive advantage, ensuring compliance with regulatory requirements, and reducing operational risks across all the business functions of companies. The corollary to this approach is the long-term convergence of Master Data Management and Big Data solutions.

In its conclusion, the study conducted with 100 organizations by PAC for Syntec Numérique⁴ identifies four major areas in which work needs to be done to assist companies with the Big Data revolution. These areas include the necessity of devising a data governance strategy involving the IS Department, the business functions, the Chief Data Officer... and defining rules for data access and security.

AN UNDERESTIMATED NEED: THE MANAGEMENT OF COMPLEMENTARY REPOSITORIES

These developments blur once and for all the boundaries between internal and external data. However, on top of them comes an increasingly acute but often underestimated problem: that of the lack of consistency and completeness of business function data and of the reference data that accompany them (the problem of managing the quality of data and complementary repositories). With a large part of data production processes being automated, one can be forgiven for thinking that the overwhelming majority of data used within enterprises is governed by quality standards and subjected to checks in this regard.

Yet, in many enterprises, it can be seen that between 2% and 5% of data are managed and collected manually⁵, outside of any IT application, to meet the requirements of the business. These requirements can be ad hoc, transitional, temporary, or, regrettably, lasting in nature. Excel files are often used to carry these data. To measure the scale of this problem, this observation needs to be considered alongside the fact that 56% of the world's enterprises state that data errors stem mainly from human errors².

Three examples:

- One company has just acquired another. Each possesses its own repositories and, while waiting for

⁴ "From data to Big Data, the expectations of business users in France"

PAC for Syntec Numérique <http://www.syntec-numerique.fr/publication/data-au-big-data-attentes-utilisateurs-metiers-france>

⁵ Internal Keyrus study and analysis of client projects

the systems to converge, the Finance Department creates an Excel file that defines the mapping between the two entities' repositories. This manually managed file is a complementary repository.

- The "products" repository is not yet integrated into the marketing department's operational system. Consequently, that department can not make the link between a campaign and the products it concerns. For each campaign, the department lists the products concerned in an Excel file and incorporates this complementary file into its system.
- In the absence of an automated process for collecting its subsidiaries' cashflow or other data, each week the head office sends an Excel template on which each entity manually enters its data. This practice can be found in the largest companies!

These complementary data and repositories managed outside of IT applications escape any rigorous control process with regard to:

- 1/ **data entry**, spreading, lead to inconsistencies in downstream applications;
- 2/ **the traceability of the data collection itself**, since with Excel it is not possible to monitor with any certainty criteria such as the dates when data is effectively updated, or to know whether all the individuals concerned have duly received the collection form;
- 3/ **how well the files are integrated with operational systems**, as cross-references can be broken by the addition/removal of an element, or simply because a name is not spelled the same way.

ENABLING THE BUSINESS FUNCTIONS AND THE IT DEPARTMENT TO BE DATA GOVERNANCE PLAYERS

The developments and practices that have just been described force enterprises to abandon data quality policies that are in silos, or limited just to the Business Intelligence information system. They militate in favour of overall data governance enabling both business functions and IT departments to contribute to enterprises' overall data quality. Keyrus knows that the absence of a simple tool for this purpose proves a major obstacle to such a policy lasting into the long term. Keyrus therefore proposes an approach backed up by a Data Governance platform shared by the IT teams and the various business function teams. This approach reaffirms and restructures the respective roles of these teams:

- IT, which ensures the overall consistency of the enterprise's extended information system, must be in control of problem issues related to the technical quality of data. Typically, it is up to IT to put in place suitable tests to ensure the precise matching up of data

between an internal system and a Cloud application, the absence of duplication, the integrity of sources, the continuity of data flows, etc.

- The Business Function teams, as users of the data, must be in control of the data's functional quality. It is true that they are better placed than the IT teams to detect inconsistencies in the indicators they use daily, or in relation to a given data history, as well as to define controls such as relevance thresholds and rules for checking.

A Data Governance platform such as **quillup**⁶ enables these two categories of players to check and monitor both the quality and consistency of the data, regardless of the nature of their sources (database, application, file, cubes, etc.), provided that the latter have been declared on the platform and the access authorizations allocated.

ELIMINATING WEAK POINTS IN THE GOVERNANCE

Data quality is never a certainty. The collection of complementary data and the creation of new repositories are inevitable in the life of an organization. Due to the absence of checks, these are also all weak points in a company-wide data governance set-up. By including in the Data Governance platform, shared by IT and the business functions, a tool specifically dedicated to managing complementary data and repositories, it is possible to eliminate the problem at its source. Instead of using Excel to collect data required to feed a downstream system, business users can do so in a rigorous and structured way using forms that automatically benefit from data entry and consistency checks. This amounts to eliminating the use of Excel files and the errors that they are liable to spread in operational and decision-making chains. To give this possibility to business function teams is to reduce the 2 to 5% of data that currently escapes any checks on quality and consistency, and to do so without having to involve the IT teams. It is also a way to limit risks of inconsistency and lost traceability during transition periods where it is necessary to create a bridge between two systems (merger, migration, etc.) and maintain cross-reference tables and temporary repositories.

AT THE END OF IT ALL: EFFICIENCY GAINS AND RENEWED CONFIDENCE IN DATA

Enterprises that have opted for this approach and which rely on the **quillup** platform for their data governance estimate that they save 30% on maintenance costs for the IT department, and that they enjoy an 80% time saving for business users, who no longer have to perform manual checks, nor call upon IT. It can be emphasized that whilst the objective is indeed to cover all the enterprise's data, the implementation of such governance must be undertaken gradually, typically by beginning with a pilot business function department.

⁶ quillup: a high performance data governance platform designed to help enterprises improve the quality of their data and make their decision-making more reliable. The platform has been developed by Keyrus through its Research & Development Center based in Israel. To find out more: <http://quillup.com/>

Key books: "Data quality and governance promoting enterprises' performance", by Laure Berti-Equille - éditions Hermès - Lavoisier

"Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program", by John Ladley - éditions Morgan Kaufmann

The essential condition for success is obviously to involve the two categories of players in the project. On the IT department side, possible objections to shared governance disappear once they are assured of retaining control over the technical quality of the data and the access rights to the various systems. Business users, contrary to what we imagine, are far from being reticent about the idea of involving themselves in managing the quality of their data. They quickly see what they have to gain from it in terms of autonomy, efficiency, and, above all, confidence in the data they use daily for operational ends or to take decisions.

In an environment in which data will increasingly stem from third-party sources not linked to the enterprise (web, social networks, connected objects, etc.), this last point is essential: the confidence of operatives in the quality of data directly determines the value that they will be able to create from those data. Beyond the immediate benefits of efficiency gains and cost reductions, it is this aspect that is no doubt the real issue at stake with overall data governance.

C.C.S.

X.D.

ABOUT THE AUTHORS

Holding a diploma in Management Computing (Paris Dauphine), **Cyril Cohen-Solal** possesses 15 years' worth of expertise in the field of Data Intelligence which has allowed him to assist leading enterprises in France and internationally with the design and implementation of their Business Intelligence strategy. Within the **Keyrus Group**, **Cyril** is also in charge of our "Keyrus Innovation Factory" accelerator, which is a bridge between innovative start-ups and major European enterprises. Alongside these activities, he has taught for 10 years at Paris Dauphine in the framework of the Master's in Business Intelligence Computing.

Xavier Dehan is Director of Big Data & Analytics Business Development within the **Keyrus Group**. He possesses broad experience in the field of IT project integration and Software Quality, acquired with major Service Companies and in the framework of the creation of Quality Insurance & Testing Pure Player companies. For over 10 years he has advised large organizations on their [Big] Data strategy for undertaking experimental and industrial projects throughout the data valorization chain.

ABOUT KEYRUS

Keyrus, creator of value in the era of Data and Digital

An international player in consulting and technologies and a specialist in Data and Digital, **Keyrus** is dedicated to helping enterprises take advantage of the Data and Digital paradigm to enhance their performance, facilitate and accelerate their transformation, and generate new drivers of growth, competitiveness, and sustainability.

Placing innovation at the heart of its strategy, **Keyrus** is developing a value proposition that is unique in the market and centred around an innovative offering founded upon a combination of three major and convergent areas of expertise:

- **Data Intelligence**

Data Science - Big Data Analytics – Business Intelligence – EIM – CPM/EPM

- **Digital Experience**

Innovation & Digital Strategy – Digital Marketing & CRM – Digital Commerce – Digital Performance – User Experience

- **Management & Transformation Consulting**

Strategy & Innovation – Digital Transformation – Performance Management – Project Support

Present in 15 countries on 4 continents, the **Keyrus** Group has 2,500 employees.

Keyrus is quoted in compartment C of the Eurolist of Euronext Paris (Compartment C/Small caps – ISIN Code: FR0004029411 – Reuters: KEYR.PA – Bloomberg: KEY:FP)

Further information at: www.keyrus.fr