The Data Value Formula

Three keys to unlock your data's full potential

UNIL | Université de Lausanne
HEC Lausanne

CDQ
COMPETENCE CENTER

CONTENT

1. INTRODUCTION

We live in an age characterized by an information explosion. According to the IDC, 59 zettabytes of data will be created, captured, copied, and consumed globally in 2020 (Reinsel, Rydning, and Gantz 2020). More data will be created over the next three years alone than over the past 30 years. The continued growth of this so-called "Global DataSphere" has also affected enterprises. They collect and store vast volumes of data in different forms and by different means – structured, unstructured, man-made, or machine-made.

This wealth of data opens up new opportunities for data-driven business, i.e. utilizing data to inform or enhance processes, decision making, and even business models. However, there is still a long way to go from simply collecting data to using data to create value for the business. Data have to be systematically managed just like any other (tangible or intangible) primary asset, such as capital assets, human resources, or patents. Only if data are "fit for purpose" and are actively used in business processes or decision-making can data create business value. This value creation process is often difficult to grasp and inherently complex.

This e-book aims to unravel the complex value creation processes using a simple model – the **data value formula**. The data value formula explains how data value can be determined and identifies three essential value drivers – **data volume**, **data quality**, and **data use**. Based on the data value formula, companies can identify and design their future value-creating steps that transform raw data into business outcomes.

To showcase the data value formula and its practical application, we illustrate **data-driven value creation in typical business scenarios**: Customer engagement, industry 4.0, and corporate performance management. We conclude by outlining the **technical capabilities** that companies must build to support value creation and show how they can be mapped to the SAP product portfolio.

2. THE DATA VALUE FORMULA

Let's start with a simple equation that helps explain data value and the factors that influence it: the data value formula.
In the digital economy, the role of data is changing. Data are turning from a secondary asset that supports business processes and decision-making to a primary asset enabling digital business strategies and business models. This **paradigm shift in the role of data** has led many companies to consider data as an asset. Nonetheless, they still struggle with determining its value.

> **"It's frustrating that companies have a better sense of the value of their office furniture than their information assets."**
> Douglas Laney, Technology Analyst at Gartner

So how can we determine data value?
Data are intangible assets – and just like any other asset, they do not create economic benefits per se, rather they only create value if used (Zechmann 2016; Laney 2017). Data value is therefore directly related to business outcomes which they help generate.

In accounting terms, data value can be determined as **value in use**. It equals the net present value of a cash flow or other benefits that data generate for a specific owner under a specific use. The quantifiable value of data can be broadly assigned into three categories:

- **generation of revenues and growth**
  (e.g. identifying new customers and market opportunities or innovating product and service offerings),
- **cost reduction and operational excellence**
  (e.g. improving business processes or optimizing asset utilization), and
- **risk mitigation**
  (e.g. responding to business, market and environmental risks or avoiding regulatory fines and penalties).

However, simply collecting more data does not necessarily create more value. Value from data is generated from a **combination of three factors** (see Figure 1): **data volume** (quantity and variety), **data quality**, and **data use**. Each factor must be actively managed by companies to generate data value.

Figure 1: Data value formula

2.1. Data volume

Creating data value starts with collecting relevant data from internal and external sources. With growing data volume – i.e. **data quantity and data variety** – the number of **potential use cases and data-driven insights** also increases. In other words, the more data organizations collect the more opportunities they have to create business value. For instance, combining internal data with external data from social media provides additional insights that help retain customers or improve up- and cross-selling.

"Over the last two years alone 90 percent of the data in the world was generated."

Bernard Marr, Author and Futurist

For many years, structured data has been at the core of data processing and has enabled enterprise-wide business process integration. However, today data-driven businesses have also started generating value from unstructured data. This unstructured data is created as user-generated content or through automatically generated digital traces.

2.2. Data quality

Collecting and storing data is only the starting point. In order to create value, **data have to be of high quality or "fit for purpose"** as well. Data quality has proven to be one of the most important barriers to unlocking data value. Poor data quality has a negative impact on its value – as reflected by the concept "garbage in, garbage out". As a rule of thumb, data at 80 percent quality means that it is only half as valuable as data at 100 percent quality (Zechmann 2016). The costs of poor data quality are often hidden, but estimated to amount to as much as 15 to 25 percent of a company's annual revenue (Redman 2017).

"Only 3 percent of companies' data meets basic quality standards."

Tadhg Nagle and David Sammon, Professors at Cork Business School; Thomas C. Redman, "Data Doc"

Whether or not data is "fit for purpose" ultimately depends on the intended use case. In the past, the purposes of data were often welldefined and linked to supporting business process and reporting. With advanced analytics, new purposes have emerged, resulting in changing expectations towards data quality. Thus, data scientists often report to spend significant time preparing and cleaning data for their purposes.

2.3. Data use

Similar to any other asset, **data can only provide value if they are used for value-creating activities.** Ultimately, if companies do not use the vast amounts of data they store and collect, it will only incur costs without creating any value. In the case of data, the logic of asset usage is even inverted. Contrary to physical assets, data increase their value when they are reused. While a machine loses value with use, a dataset is non-rival; i.e., it can be used by different persons at the same time, and increases in value the more it is shared and reused. In line with this approach, a recent study highlights that companies with established data-sharing practices have a higher ability to innovate (Ransbotham and Kiron 2017).

"Less than half of a company's structured data is actively used in making decisions, and less than 1 percent of all unstructured data is analyzed at all."

Leandro DalleMule, CDO AIG, and Thomas Davenport, Professor at Babson College and MIT

To monetize their data, companies should actively search for new ways of using data to generate business outcomes:

- enhance or improve business processes by means of data-driven insights,
- improve existing products and services by "data wrapping", and
- provide stand-alone data services.

3. KEYS TO UNLOCK BUSINESS VALUE FROM DATA

Value from data is generated by managing volume, quality, and using data. None of these steps work in isolation. Collecting data does not create value; neither does investing in and maintaining a small subset of high-quality data but never using it. The **value-adding steps** needed for this transformation must work in alignment and form the so-called **data value chain** (see Figure 2).

3.1. Volume: Source data in the right quantity and variety

While data volume is an important value driver, it is also a conundrum. In practice, most enterprise data remain "hidden" in the large number of systems and databases owned by the enterprise. This phenomenon is also called **dark data**. Dark data generate costs (e.g. for storage, backup or administration), but do not create value. Enterprises that do not know where they store their data or if their data is being used will find it difficult to generate business value.

This implies that companies must **acquire and create data in a transparent way** to ensure that the data are collected in the **right quantity and variety** for the relevant business purposes. For unstructured data, like documents or behavior data, important information should be documented through metadata. For structured data, data creation should adhere to defined semantics and rules so that data users can trust that the data are correct, accurate, and consistent.

3.2. Quality: Make data "fit for purpose" and manage the data life-cycle

The acquired and created data are then transformed into **data assets** through systematic organization and management, with the goal of building a solid **data foundation.** Data assets are bundles of all the different data points that can be related to core business objects (such as customers, products, employees, or company branches) and are the basis of value creation. These **data assets have a specific lifecycle – from creation to retirement** – and should be managed just like any other (tangible or intangible) primary asset. Similar to production and maintenance processes for physical assets, data and their lifecycles must be actively managed. This includes first and foremost the addition of semantics and quality to the data to ensure that data assets are "fit for purpose".

Importantly, **interrelated data from different sources must be connected** and a comprehensive view on relevant business objects created. In order to do so, key data objects need to be uniquely identified and allow for connection. An important aspect of data management is the monitoring of data's quality and possible remediation. Data must also be retired at their "end of life", as defined by internal data retention rules or external regulations. Since enterprises have to comply with an increasing number of data protection regulations, data management ensures that restrictions on certain categories of data (e.g. personal identifiable data) are understood and respected, and that data are deleted if requested.

3.3. Use: Actively use data assets to produce business outcomes

Finally, data consumption comprises the **different ways of using data to deliver business outcomes.** Companies achieve business value from data by integrating data-driven ways of working into their operational business processes and decision-making. They increasingly enrich existing products and services with data, or monetize data themselves. In order to leverage these data monetization opportunities, data need to be findable and accessible by those who need them. In addition, the silos between analyzing data and actively using these data have to be removed. Generating higher data value can only happen when the insights and findings are directly embedded into execution.

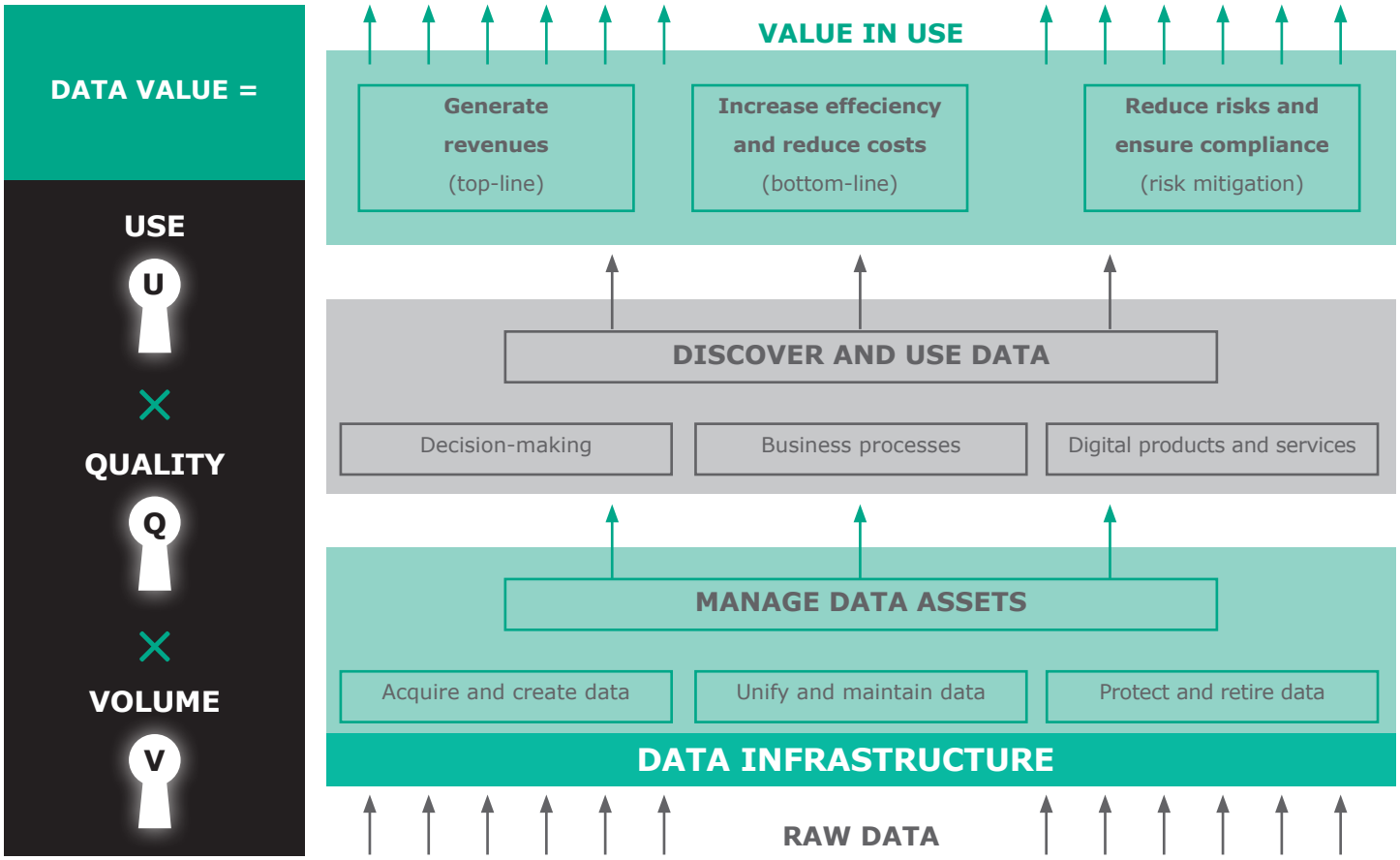

Figure 2: Data value chain

4. PUTTING THE DATA VALUE FORMULA INTO PRACTICE

4.1. Customer engagement

Knowing your customer ("KYC") and understanding customer behavior are essential capabilities for companies in an increasingly digital world. Digital traces from customers give enterprises the ability to not only understand customers' preferences and their behavior, but also engage with customers in new ways.

Data value: Data ultimately help maintain (improving customer retention) and expand (increasing conversion rates) a company's customer base, while also increasing customer life-time value (increasing cross-selling and upselling).

DATA VALUE =

- Increasing customer life-time value through cross-selling and upselling opportunities and personalized offerings
- Improving customer retention to avoid churn
- Increasing conversion rates to expand active customer base and improve sales efficiency

USE

U

×

- Analyzing customer journeys to predict and prevent churn or identify high probability leads and prospects
- Analyzing customer behavior to derive real-time segmentation and improve customer targeting (e.g. campaigns)
- Analyzing customer preferences and interests to provide recommendation or individualize products and services

QUALITY

Q

×

- Connecting data across the silos of different channels and source systems (Web, CRM, ERP, …) – and organizational silos (marketing, sales, service)
- Unifying and integrating data through unique identifiers and master data for customers
- Classifying data according to their sensitivity
- Cleaning data (reactive or proactive)

VOLUME

V

Collecting data across all touch points and channels:

- Direct contacts (emails, phone, meetings, etc.)
- Automated tracking of customer behavior ("digital traces") on websites
- Customer transactions such as orders, invoices, complaints, etc.

Sourcing relevant external data:

- Social media profiles
- Competitive analysis and market statistics

Figure 3: Data value chain "Customer engagement"

Data volume: First of all, companies should make sure to source the right quantity and variety of data to achieve a 360° view on their customers. This implies collecting data from a large variety of systems across different touch points and channels. These sources include direct contacts (e-mails, phone, meetings, etc.), automated tracking of customer behavior ("digital traces") on websites and online channels, and customer transactions, such as orders, invoices, complaints, etc. External data adds market insights, for instance from market research reports or statistics, or on a more detailed level, from social media profiles or news.

Data quality: The collected data must then be made "fit for purpose". Unification and integration are key for creating value from this quantity and variety of data and breaking down organizational silos in marketing, sales, and service. Once data have been brought together, unique identifiers and customer master data help in gaining a unified 360° perspective over all sources. Importantly, unifying customer data means resolving similar entities on varying attribute names and connecting multiple sources based on common references. This step usually includes cleaning data with reactive and proactive approaches. It must also be noted that any processing of identifiable information must be done in alignment with privacy regulations.

Data use: Having this unified and integrated view on customers enables multiple usage scenarios and personalized engagement with customers. Analyzing customer journeys can prevent churn while helping to identify high probability leads and prospects. Identifying customer segments and tracing their behavior (e.g. click-streams) means customers can be targeted in real-time (e.g. campaigns). Finally, analyzing customer preferences and interests makes it possible to provide recommendations or customize complete products and services.

4.2. Industry 4.0

With the fourth industrial revolution, manufacturing companies increasingly use **smart, connected technologies to automate traditional practices of producing goods**. A key challenge is to ensure that production processes in these companies are cost-effective and continue uninterruptedly. Manufacturing companies build "**digital twins**" of their assets to provide a real-time representation of their physical assets.

Data value: The "digital twins" can then help to reduce (unplanned) downtimes and mitigate risks while improving asset utilization and operational efficiency. The increasing availability of fine-grained data from sensors also enables completely new business models like "pay per use" or metered services.

Data volume: Industry 4.0 scenarios leverage the availability of fine-grained data from various sources and connect them to transaction data.

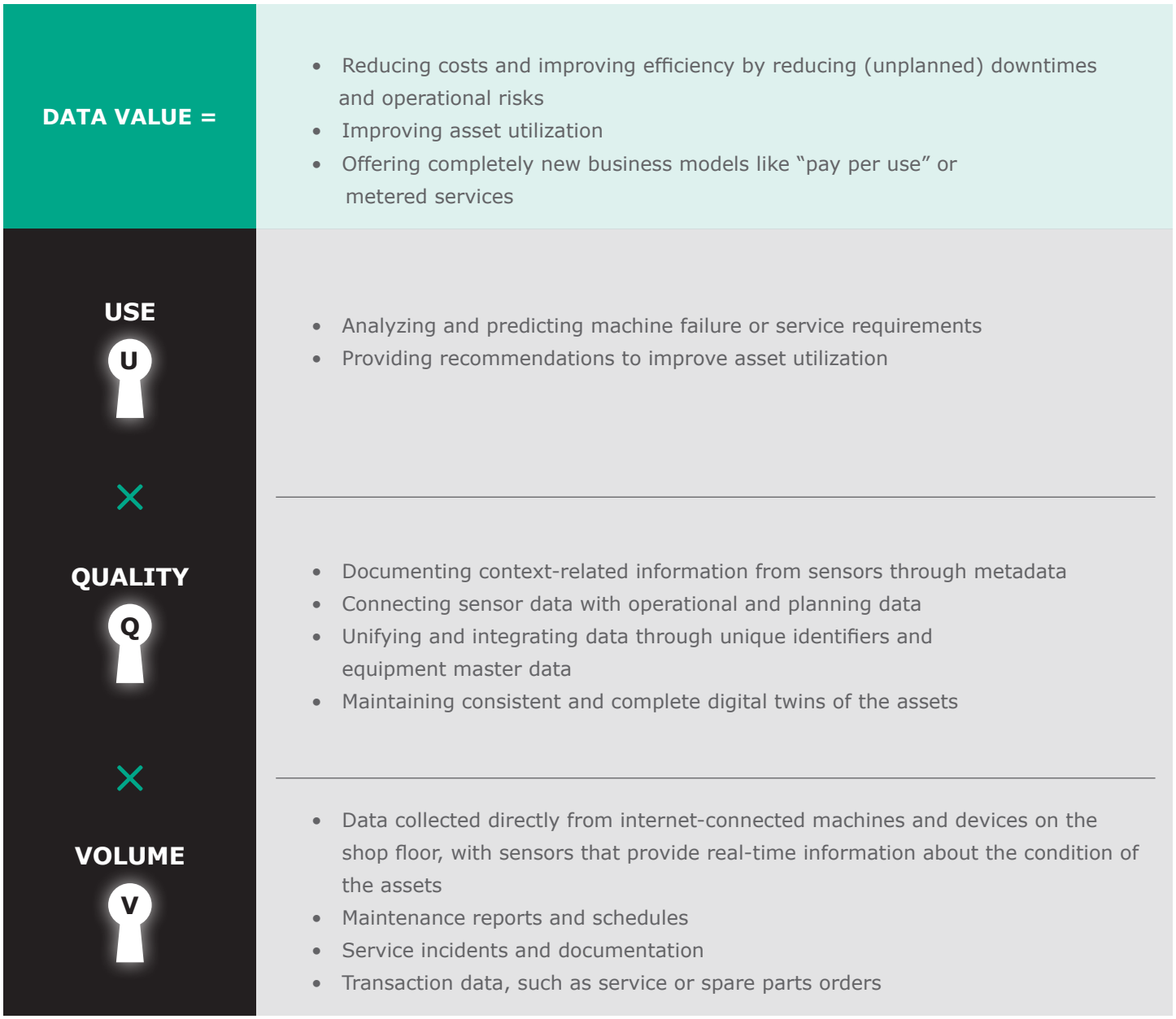

Figure 4: Data value chain "Industry 4.0"

For instance, internet-connected machines and devices on the shop floor provide real-time information about the assets' condition via sensors, such as data from vibration, acoustic, ultrasonic, temperature, energy consumption, oil analysis, and thermal imaging sensors. Furthermore, maintenance reports, schedules, and service incidents document the status of assets and key events over time. Transactional data such as service or spare parts orders also give further indications about when and how an asset should be repaired.

Data quality: Before data can help to improve maintenance operations, the collected data must first be unified and integrated, typically in a data lake. Next, data from sensors need to be connected with operational and planning data. To establish this connection, unique identifiers and equipment master data need to be established. Finally, real-time data from the sensors and the metadata with contextual information is ingested.

Data use: Once this unified view of the equipment has been created, data can be analyzed to predict machine failure or service requirements.

4.3. Corporate performance

As business environments become increasingly dynamic, enterprises must foster **agility to react to unforeseen events** while quickly adopting innovative technologies and new business models.

Data value: To deal with these changes effectively, companies need to rigorously control their investments, align strategies and business results more often, and look to foresee risks. Data can facilitate these tasks and help reduce financial risks, optimize costs, and drive revenues.

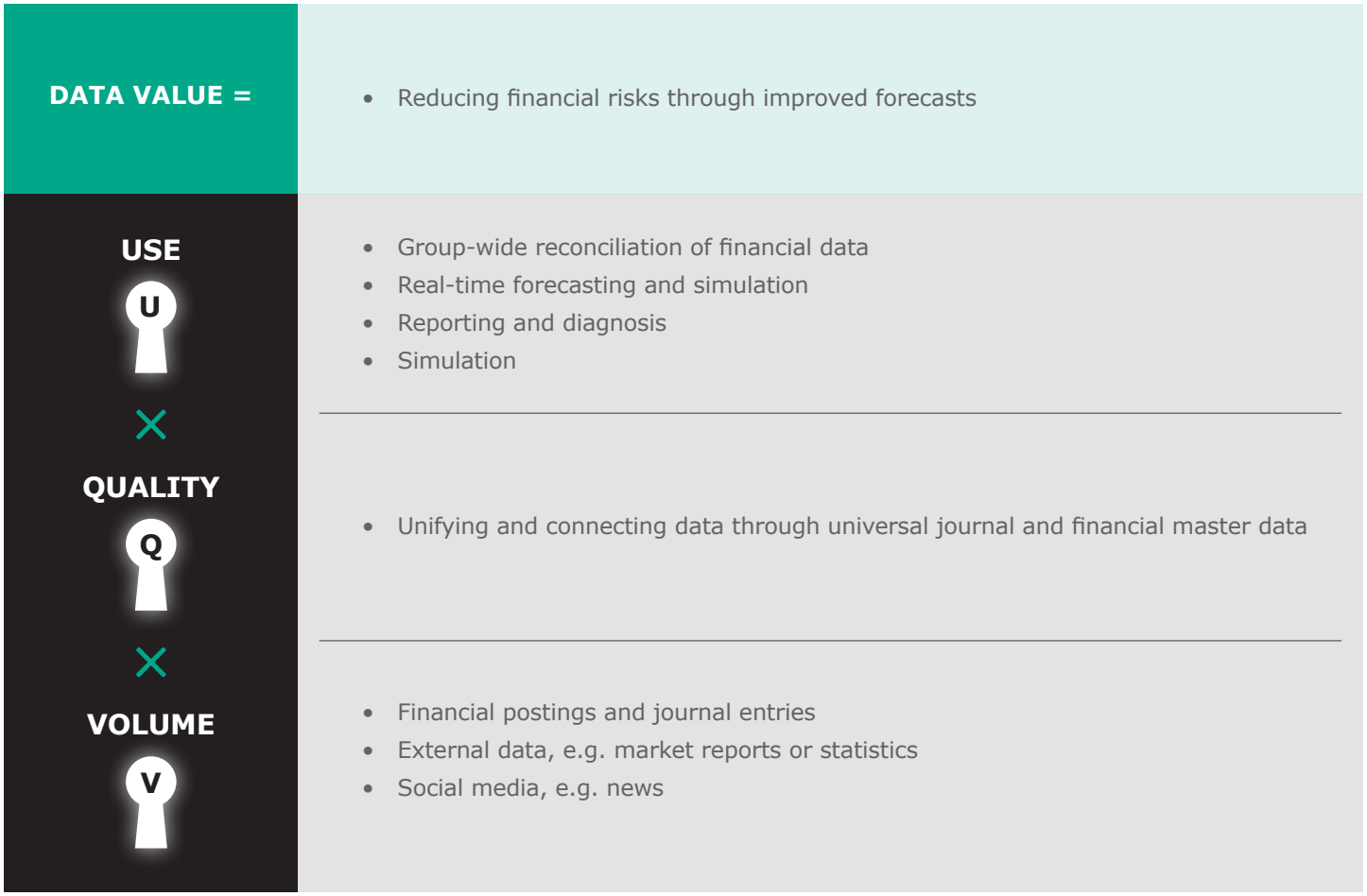

Figure 5: Data value chain "Corporate performance"

Data volume: In practice, internal data regarding investments and financial status are leveraged by corporate performance management at an enterprise level. Financial transactions like financial postings and journal entries are important for monitoring a company's financial status. These data can be complemented by external economic data captured in market reports or news articles, for instance.

Data quality: For global companies, harmonized financial master data and a universal journal help provide a uniform view of these different sources and maintain a consistent general ledger.

Data use: On top of that, reconciliations can be run efficiently, and planning processes can be facilitated through forecasts and simulations. In addition to internal data, external data concerning markets, competitors, and general trends allow a company to assess their performance against its direct business environment. This assessment can help a company adjust its strategy and roadmaps accordingly. In times of uncertainty, real-time news monitoring helps detect relevant events and disasters that might impact supply chains and company performance.

5. SUPPORTING THE DATA VALUE FORMULA WITH TECHNICAL CAPABILITIES

Any company that wants to turn data into business value must incorporate three core technical capabilities:

- infrastructure capabilities for collecting, storing, and processing data,
- capabilities to manage data assets and making data "fit for purpose",
- data consumption capabilities that use data and embedded analytics to improve business processes, enhance existing products, and deliver dedicated data products / services.

Let's break down each capability.

5.1.Capabilities for storing and processing data (data infrastructure)

The ever-increasing quantity, variety, and velocity of data means enterprises need scalable data infrastructure capabilities to store and process large amounts of data in different formats. Up till now, most existing data infrastructure has focused on structured data in relational formats. However, this infrastructure needs assistance to cope with the increasing volume of unstructured data collected by sensors and user-generated content. In addition, it must provide integration and efficient access to the increasing number of databases and source systems, while reducing the amount of copy procedures.

KEY QUESTION	• How to **store and process** the increasing **data quantity** and **variety**?
EXEMPLARY CAPABILITIES	• **Storage** (relational and no-sql) • **Transactional** and **analytical** (batch and stream) processing • **Ingestion** and **transformation** (ETL/ELT) • **Staging / landing** • **Querying** • **Virtual data access**
STAKEHOLDERS	• IT managers • Database and infrastructure experts

Table 1: Data infrastructure capabilities

Companies that establish these capabilities are able to store and process structured data as well as the increasing amount of unstructured data.

5.2. Capabilities for managing data assets (data lifecycle and quality)

Companies must be able to manage data assets along their entire lifecycle and eventually make data "fit for purpose". The required capabilities include managing metadata, monitoring data quality, unifying data across sources, identifying sensitive data, and managing data's lifecycle. Data quality management comprises data profiling and quality measurement, as well as data enrichment and cleaning. Finally, technical capabilities are needed to manage the lifecycle of the core master data objects from creation to maintenance and retirement. An important capability relates to data protection and the handling of sensitive data in alignment with internal or external regulations.

KEY QUESTION	• How to make data **fit for purpose**?
EXEMPLARY CAPABILITIES	• **Metadata** and **data modelling** • **Data quality** management • **Master** and **reference data management** • **Lifecycle management** • **Data protection** (classification of sensitive data, consent management, etc.) • **Archiving**
STAKEHOLDERS	• Data managers • Data protection and compliance experts • Enterprise and data architects

Table 2: Data lifecycle and quality capabilities

These capabilities help companies make data "fit for purpose" for the relevant business scenarios.

5.3. Capabilities for data discovery and use

Data is turned into business value or outcomes with a variety of applications. Analytical applications are used to explore and analyze data as well as experiment with advanced analytics models. Today, analytics are increasingly directly embedded in operational applications to support higher levels of automation, context-aware processing and decision making in real-time.

Beyond analytical and operational applications, an additional category of dedicated data applications is emerging: Data catalogs, data marketplaces or data platforms mediate between data supply and demand. They help the increasing number of data consumers find, understand, use, and collaborate on data.

KEY QUESTION	• How to make the best **use of data**?
EXEMPLARY CAPABILITIES	• **Context-aware workflows** / **business processes** • **Robotic process automation** • **Embedded analytics** in operational systems (Insights to Action) • **Planning** and **simulation** • **Self-service analytics** • **Data discovery**
STAKEHOLDERS	• Data consumers • Data experts (data scientists, etc.)

Table 3: Data discovery and use capabilities

Companies that incorporate these capabilities are able to monetize their data. They achieve monetization by integrating data-driven ways of working into their operational business processes and decision-making as well as innovating their business models.

6. HOW SAP APPLIES THE DATA VALUE FORMULA

The data value formula underpins SAP's solution strategy. The SAP Business Technology Platform is a portfolio of integrated solutions that turn data into business value. It tightly integrates individual SAP solutions that focus on each step of the formula – volume, quality, and use – to unlock your data's full potential. SAP Business Technology Platform includes database and data management, application development, integration, analytics, and intelligent technologies – from on-premise to the cloud.

Different products in the platform focus on solving each part of the formula. SAP HANA Cloud for volume, SAP Data Warehouse Cloud and SAP Data Intelligence Cloud for quality, and SAP Analytics Cloud, SAP Integration Suite, and SAP Extension Suite for use. Although these solutions can be used individually, they seamlessly integrate to unlock the full value of data.

Data volume

SAP HANA Cloud is a multi-cloud database as a service (DBaaS) that stores and tiers data to ensure that high-value data are always accessible. It does all of this with high speed and scalability. SAP HANA Cloud can transfer hot data from in-memory to warm in the native storage extension or cold in the data lake. Data stored in memory are available in real-time with no latency. As data become less of a priority, it can be moved to reduce cost. SAP HANA Cloud can also power the development of new applications or it can connect to an organization's existing applications.

Data quality

SAP Data Intelligence Cloud covers the integration, orchestration, and automation of data. It offers robust pipelining and data cataloging to help manage data of any volume and transform it for quality.

SAP Data Warehouse Cloud is an enterprise-grade data warehouse in the cloud built for IT and line of business users. It ensures data quality by centralizing data sources and providing one source of truth. SAP Data Warehouse Cloud offers one robust data layer with data models that span multiple data sources. Data are replicated or virtualized to provide organizations with trusted and up-to-date data.

SAP Data Warehouse Cloud's Spaces – virtual work environments – enable administrators to change data access or allocate computing and storage resources for specific employees or teams. It empowers business users to connect data, model, and visualize themselves with premade industry-specific templates and native integration with SAP Analytics Cloud.

Data use

SAP Analytics Cloud combines three critical domains of BI, planning, and augmented analytics to help organizations use data and make end-to-end decisions confidently. This analytical cloud solution can discover deep insights and empower informed decision-making through machine learning and built-in artificial intelligence. SAP Analytics Cloud seamlessly integrates with data to simplify an organization's analytics landscape. Connecting to data from multiple sources leads to a full picture of the business and better-informed decisions.

SAP Integration Suite and SAP Extension Suite brings powerful managed development and integration capabilities to individually shape the flow and use of data. With SAP Integration and SAP Extension Suite services, tools, and APIs IT teams can easily extend functionality of business solutions and business processes and connect with any SAP and non-SAP system or data source.

For an individual and optimized digital experience, the platform brings built-in digital experience and business service capabilities that allow the creation of efficient, machine learning supported, and engaging user interfaces, interactive chatbots, mobile apps, automated business processes, intelligent documents and ticket classification, and predefined workflows. Further details about all services can be found on the SAP Discovery Center.

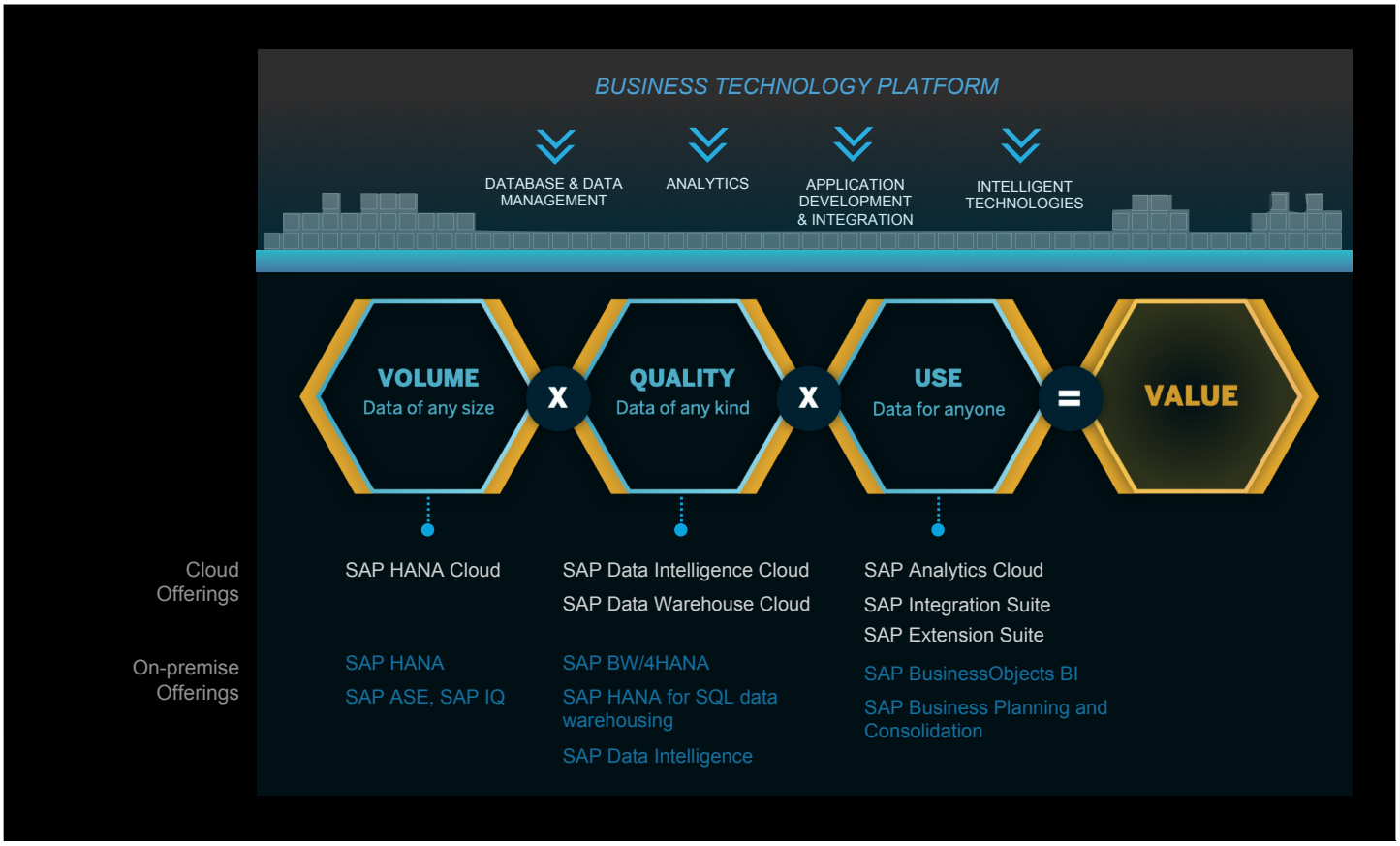

Figure 6: SAP Business Technology Platform

7. REFERENCES

Laney, Douglas B. (2017) Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage. Routledge.

Ransbotham, Sam and Kiron, David (2017) Analytics as a Source of Business Innovation. MIT Sloan Management Review. https://sloanreview.mit.edu/projects/analytics-as-a-source-of-business-innovation/, accessed November 3, 2020.

Redman, Thomas C. (2017) Seizing Opportunity in Data Quality. MIT Sloan Management Review. https://sloanreview.mit.edu/article/seizing-opportunity-in-data-quality/, accessed November 3, 2020.

Reinsel, David, Rydning, John, and Gantz, John F. (2020) Worldwide Global DataSphere Forecast, 2020–2024: The COVID-19 Data Bump and the Future of Data Growth. IDC. https://www.idc.com/getdoc.jsp?containerId=US44797920, accessed November 3, 2020.

Zechmann, Andreas (2016) Data Valuation. Working Report. Competence Center Corporate Data Quality (CC CDQ). https://www.cc-cdq.ch/publication/assessing-the-economic-value-of-data-assets, accessed November 3, 2020.

Resource	Link
Competence Center Corporate Data Quality	www.cc-cdq.ch
CC CDQ Data Valuation	www.cc-cdq.ch/data-valuation
CC CDQ Data Strategy	www.cc-cdq.ch/data-strategy
CC CDQ Using AI for data management	www.cc-cdq.ch/Machine-Learning-for-Improving-Data-Quality
Business Technology Platform	www.sap.com/products/business-technology-platform.html
SAP HANA Cloud	www.sap.com/products/hana/cloud.html
SAP Data Intelligence Cloud	www.sap.com/products/data-intelligence.html
SAP Data Warehouse Cloud	www.sap.com/products/data-warehouse-cloud.html
SAP Analytics Cloud	www.sap.com/products/cloud-analytics.html
SAP Integration Suite	www.sap.com/products/integration-suite.html
SAP Extension Suite	www.sap.com/products/extension-suite.html
SAP Discovery Center	https://discovery-center.cloud.sap/

ABOUT THE COMPETENCE CENTER CORPORATE DATA QUALITY

The Competence Center Corporate Data Quality (CC CDQ) is a research consortium and expert community in the field of data management. CC CDQ's mission is to **support companies in managing data as an asset** by developing and transferring innovative approaches into everyday business practice.

Established in 2006, the work of the CC CDQ is carried out in **close collaboration with more than 15 European enterprises from various industries** (see Figure 7). It is driven by the requirements of innovative business models inspired by digitalization, global market presence, worldwide business process harmonization, industrialized services, and customized operations. Success in these areas relies on consistent, accurate, complete, highly available, secure, and timely corporate data resources.

The CC CDQ activities comprise three pillars:

- **Networking** with data management experts in the CC CDQ community through workshops and informal exchanges.
- **Research and co-innovation** to develop innovative, practice-proven solutions and approaches for enterprise-wide data management.
- **Knowledge sharing and good practices** in the CC CDQ Knowledge Base containing use cases and company presentations, concepts, methods and tools, research papers and working reports.

PARTNER COMPANIES OF THE CC CDQ

Figure 7: Current member companies of the CC CDQ

AUTHORS

Prof. Dr. Christine Legner
Professor of Information Systems
Faculty of Business and Economics (HEC)
University of Lausanne

Academic Director
Competence Center Corporate Data Quality (CC CDQ)

Martin Fadler
PhD Candidate in Information Systems
Faculty of Business and Economics (HEC)
University of Lausanne

Research Associate
Competence Center Corporate Data Quality (CC CDQ)

Title: © AdobeStock | December 2020

UNIL | Université de Lausanne
HEC Lausanne

CDQ COMPETENCE CENTER

This e-book was created in collaboration with SAP SE.

CDQ AG | Lukasstrasse 4, CH 9008 St. Gallen | www.cdq.ch

Faculty of Business and Economics (HEC Lausanne) | CH 1015 Lausanne | www.hec.unil.ch