Revista de Contabilidad y Dirección
Vol. 31 año 2020, pp. 75-90

Information Management – a holistic approach

JENS ROPERS
CA Controller Akademie, Wörthsee, Germany

Reception date: 14-7-2020
Acceptance date: 12-10-2020

ABSTRACT

In the context of digitization initiatives and projects for corporate management the attention is often exclusively on Business Intelligence, *Big Data* and Advanced Analytics. Here, the initial focus is on the extraction, utilization and storage of data. Those can then be analysed by means of a classic or advanced statistical methods and/or machine learning algorithms in order to gain insights to support management decisions. Despite all fascination for analytics and technology, further steps should not be overlooked. In an increasingly dynamic environment, it is reasonable to constantly review the company's management model and adapt it, if necessary. In addition to that it is of importance to prepare and use the information obtained in a way that is understandable for the management and to provide decision-makers with the necessary data competence to be able to follow the path from data to decision. In this way the important bridge from obtaining and analysing data to management decisions is being built.

Classification JEL: MO, Z0

KEYWORDS

VUCA-World, Ambidexterity, Data Architecture, Advanced Analytics, Self-Service Business Intelligence, Data Literacy.

RESUMEN

En el contexto de iniciativas y proyectos de digitalización para la gestión empresarial, la atención suele centrarse exclusivamente en Business Intelligence, *Big Data* y Advanced Analytics. El enfoque inicial se centra en la extracción, utilización y almacenamiento de datos. A continuación, estos se pueden analizar mediante métodos estadísticos clásicos o avanzados y/o algoritmos de aprendizaje automático para obtener información que respalde las decisiones de gestión. A pesar de toda la fascinación por el análisis y la tecnología, no se deben pasar por alto los pasos adicionales. En un entorno cada vez más dinámico, es razonable revisar constantemente el modelo de gestión de la empresa y adaptarlo, si es necesario. Además de esto, es importante preparar y utilizar la información obtenida de una manera que sea comprensible para la gerencia y proporcionar a los tomadores de decisiones la competencia en datos necesaria para poder seguir el camino desde los datos hasta la decisión. De esta manera se está construyendo el importante puente entre la obtención y el análisis de datos y las decisiones de gestión.

Clasificación JEL: MO, Z0

PALABRAS CLAVE

Mundo-VUCA, Ambidestreza, Arquitectura de datos, Análisis avanzado, Inteligencia empresarial de autoservicio, Alfabetización de datos.

---

1. Dynamics and Uncertainty

The dynamics of today's world demand agile decision-making. Therefore we need methodical and technological requirements as well as the necessary skills!

The acronym VUCA (Harvard Business Review, Feb 2014) was coined within the American military to describe the greatly changed situation on the modern battlefield. Instead of clearly demarcated, ordered fronts, more and more confusing conflicts arose. Fortunately, the activities of a company are not fully comparable with military conflicts. Nevertheless, there are a number of examples which show that more and more companies today are operating in a "VUCA world". The V of VUCA stands for volatility, an instability with a high fluctuation margin, for example in share prices and commodity prices. U as in uncertainty stands for the uncertainty caused by surprising, unpredictable situations and challenges, such as changes in legislation or the diesel scandal. C as in complexity describes the complexity caused by incomprehensible cause-and-effect relationships, for example in the price development of cryptocurrencies. And A for ambiguity describes the ambiguity of information and observations, such as fake news. In addition, new business models, products, and services are being created at breakneck speed, especially as a result of the digital transformation. Some disappear just as quickly; others stabilize or change. What seemed to make sense yesterday is obsolete today. The Covid-19 crisis has drastically exacerbated an already challenging situation. In an impressive way it became clear how difficult it is to find the right key figures and make decisions in a highly dynamic and complex world. In many countries, very different approaches were decided on how to proceed in the crisis. There were also widely differing key figures that were considered relevant. If robust cause-and-effect relationships are becoming increasingly difficult to identify and even the objective itself is constantly being questioned, an uncertainty arises that cannot be solved with large amounts of data alone. Even with the use of complex algorithms, the future will not be validly predictable. Rather, it is necessary on the one hand to get involved with the black box of an algorithm and on the other hand to critically reflect on such a result and put it into context with experiences and current influences. In concrete terms, this means, for example, that management and controlling will in future have to increasingly engage in thinking in bandwidths and discuss the probability of the occurrence of different scenarios. The parameters that influence the probability of occurrence can change rapidly in a highly dynamic world, so that decisions must be regularly reviewed and adjusted as necessary.

But even when the objectives have been clarified and the management model has been developed and backed up with data, there is still one last important step to take: The decision makers must be provided with the necessary information to be able to make decisions within the company. This is done by the *reporting* system, which presents the essential information in an understandable way. This requires interlocking cooperation between IT, controlling, and management. IT is responsible for the development of system landscapes and data architectures that are capable of meeting the rapidly changing requirements of controllers, business units, and management. Controllers and/or data scientists analyze the data provided by the IT systems. Subsequently, pure statistics are generated, which do not contain any messages and whose interpretation must be carried out by the management itself. Alternatively, reports can be generated. These differ in that they are enhanced for management by the addition of messages and thus facilitate decision-making (Hichert and Faisst, 2017: V-VII). We summarize all these aspects under the generic term Information Management (Fig. 1). It is imperative that there is intensive interaction between controllers, management, and IT. In addition, new roles are emerging that may also be of interest to controllers in the future.

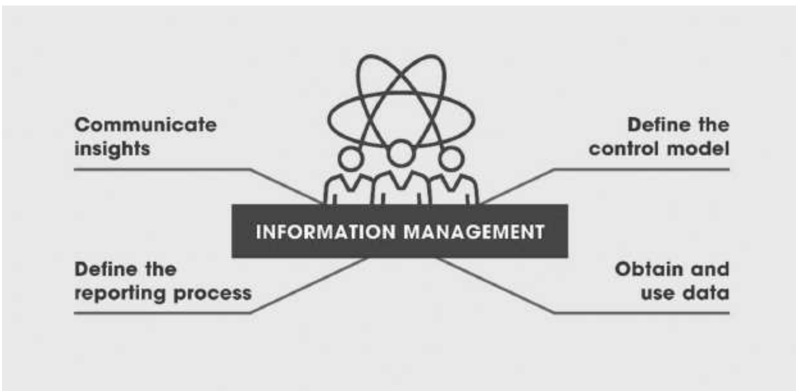

Figure 1. Holistic view of Information Management.

2. Define the control model

The basis for the control model is the corporate strategy. Strategic and operational goals are developed on this basis. In addition to aligning the company to specific markets and customers, and thus creating future potential, the focus is primarily on achieving the financial goals of the current and coming fiscal year. In order to achieve this, corporate processes have been optimized for decades, documented with key figures, and managed in line with the targets. The focus has always been on efficiency in the design of corporate processes. The aim was to meet deadlines while maintaining high quality at the lowest possible cost. This pattern is of course still valid today. Only no longer exclusively. “Ambidextrous” companies have recognized that the increasing dynamics of the VUCA world cannot be reflected in standardized processes alone.

In addition, a sensible approach to dynamics is needed. Dr. Gerhard Wohland describes companies that are prepared for the challenges of the future as dynamically robust (Wohland and Wiemeyer, 2012). The robust part is based on knowledge and experience, which are incorporated into processes via rules and methods. If the processes are no longer running efficiently, they are analyzed for weak points. A process improvement is then usually based on benchmarks, best practice examples, or internal measures. This is based on the assumption that the environment that determines the process is at least almost completely recorded and understood and that a new target process can be defined. The problem, however, is that in our volatile and complex world, the speed of change has increased rapidly, with the result that target processes that have just been defined no longer come to rest. Permanent adjustments and adaptations have to be made, often resulting in chaos. In reality, this is an attempt to force increasing dynamics into standard processes. Exactly this leads to the fact that the company neither has efficient processes nor can deal with dynamics. The essential success factor for dealing with dynamics is the realization that we can no longer force every task into clearly defined process steps, work instructions, and organizational responsibilities. Rather, we need a jointly supported target state that is worked toward. In doing so, the employees orientate themselves on principles which become guidelines in order to make their own scope for decision-making comprehensible. When solving problems, tools are applied in a variety of ways or even combined, and certain predefined methods of resolution are not used. And finally, work is more project-like in interdisciplinary teams than in functionally organized departments to pursue silo thinking and isolated objectives. "The successful thinking traditions for sluggish markets today are not the solution, but the problem!", says Wohland (www.dynamikrobust.com). This also applies to the controllers, of course. Experience has shown that the dynamically robust world for controllers is most evident in the control model and the resulting data requirements and *reporting*. For their own control, the specialist departments always need different key figures and indicators. Control variables have to be redefined on a regular basis, or ad hoc reports have to be created for a short period of time, which disappear again shortly afterwards. Here, an agile approach and the use of OKR (Doerr, 2018) in addition to classic key figure systems can be a sensible alternative. In addition to the company's objectives and the control variables that help to achieve them, the respective dimensions, the comparisons necessary for controlling (plan/actual, target/forecast, etc.) have to be redefined again and again and the necessary data have to be obtained.

3. Obtain and use the necessary data

In all probability, most of the classic financial key figures will remain unchanged for a long time in the future and their robust management ratios will be stored in a structured manner in the data warehouse of a Business Intelligence (BI) system and integrated into reports from there. For those metrics that change more frequently, the basic data necessary for the calculation must be captured anew in each case, which demands greater flexibility from the BI systems. The more decentralized the control, the more individual is the data requirement. This is often a field of tension between flexibility and data governance. For their own control, the departments need constantly changing data and information. IT pays attention to the single source of truth and data consistency and therefore insists on certain procedures and rules. Thus, controllers who serve business partners who are responsible for business departments often find themselves in a dilemma, because both business department and IT are right from their respective perspectives. Most of the time, the subject of *reporting* does not remain with three parties. Because controller departments are becoming more and more specialized, for example in the direction of data science, and because software providers often have to be involved, four to five departments are often involved in generating information. The overview in Fig. 2 should help to bring together the most important aspects of data procurement and use. The overview is always based on the previously defined control model, or specific questions (use cases) that need to be answered. A pure collection and storage of data without a concrete objective is less recommended.

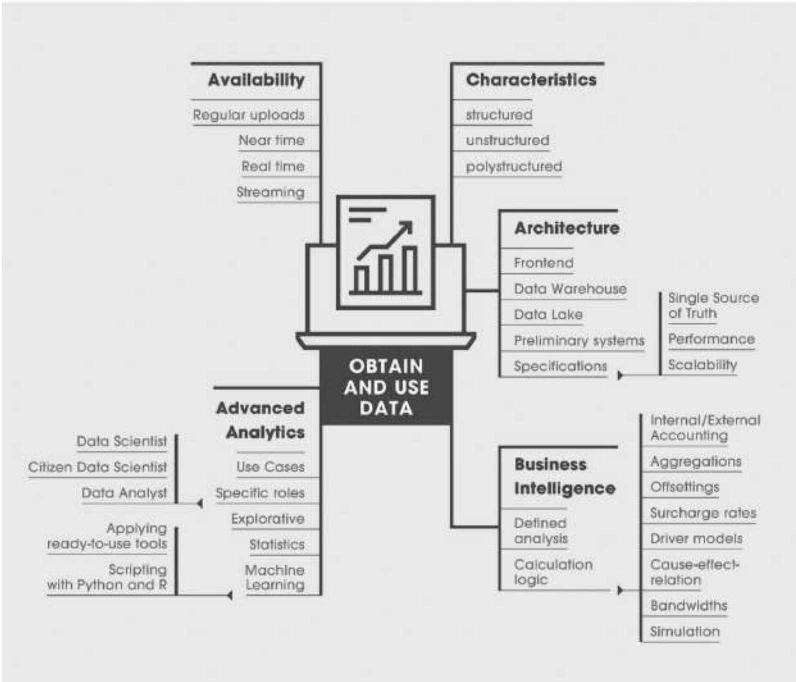

Figure 2. Obtain and use data.

It is advisable to determine the expected future information and data requirements as part of a BI and *Big Data* strategy. On this basis, the data architecture, organizational requirements, and the necessary roles can be determined. If you initially only need regular batch uploads of structured data that is aggregated into key figures using clear calculation rules and then combined into monthly reports in a frontend, a corresponding data warehouse with a connection to the upstream systems and a frontend is sufficient. If polystructured data (e.g. text, images, figures) from a wide variety of national companies and social media sources are to be collected in real time and analyzed with machine learning, a Data Lake architecture (Gorelik, 2019) with corresponding analysis tools and/or data science is more suitable.

Business Intelligence

In the first step, it is useful to take a closer look at the terms Business Intelligence (BI) and *Big Data*. BI comprises the collection, storage, and processing of structured data to support statistics, reports, and analyses in planning and in the context of decision-making processes. As a rule, small to medium-sized data volumes are loaded from previous systems once or several times a day. The BI Data Warehouse architecture shown in Figure 3 is very well suited for this purpose. It is very important to completely decouple the data warehouse area from the source systems AND the frontend and especially NOT to access the source systems directly with frontends! Only in this way is it possible to connect new ERP systems without any problems and, if necessary, to change to another frontend. Not to mention the maintenance of data consistency.

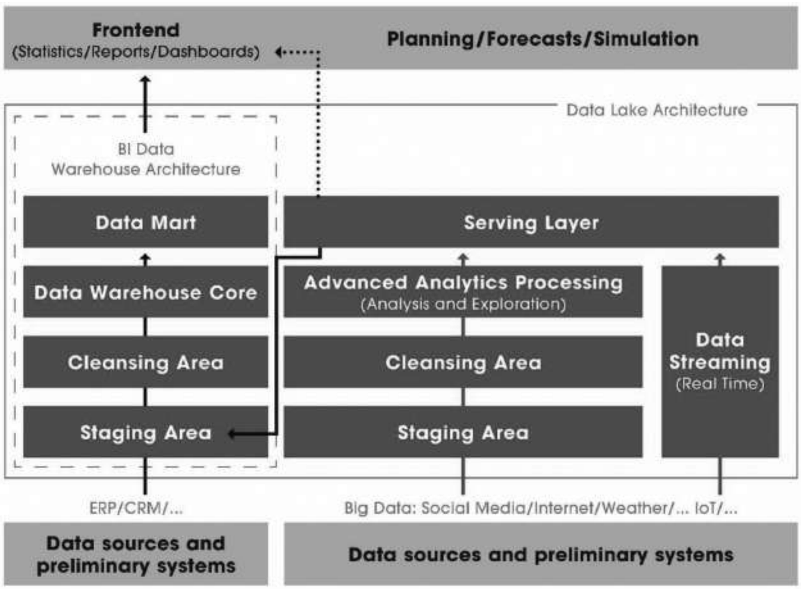

Figure 3. Overview of the BI/*Big Data* data architectures (based on QUNIS GmbH).

Let's look at the diagram starting from the bottom. The data from the preliminary systems are loaded into the staging area. Here, the data are entered either in the source tables or in individual tables. It is crucial for the long-term success of a BI initiative that the path via the staging area is the only possible way into the data warehouse and therefore the only way into the frontend and *reporting*. There is no other way to permanently maintain a "single source of truth". In the next layer, the cleansing area, data changes to previous runs are determined and consistency checks are performed. If necessary, data are cleansed and/or completed. The data are transferred into the structure of the data warehouse target model. For example, date formats etc. are adapted to the required standard. In the Data Warehouse Core, data changes are historicized, data refinements are carried out, additional calculations with company-wide character are made, and special functions such as time series calculations are performed. Finally, in the data mart, relational or multidimensional structures are created, which are required to provide data models for later front end use. In addition, department-specific transformations, summarizations, or additional calculations are generated. The frontend uses the data mart. This has no direct connection to the front end systems, but only accesses the data mart. There should also be no further data retention in Excel sheets or other tools. The structured data in the data mart is the "Single Source of Truth"!

The term *Big Data* initially only refers to the fact that data are available in ever-increasing quantities and variants and can be processed at ever increasing speed, and now also include methods and technologies for the highly scalable acquisition, storage, and analysis of even unstructured and polystructured data. By this we mean data that –mostly due to its origin– is available in completely different formats. This can be RFID radio chips and GPS data, (text) notes from call centers or social media, images and video clips from the Internet, machine data from production, or log data from your website. These are regularly loaded into a staging area. Regularly means here, for example, loading batches several times an hour up to near-time data, which are permanently loaded with a time delay of only a few minutes or even seconds. The more sources are used and the more continuously the data are loaded, the higher the volume of data to be processed (which is available in the most diverse structures). This is the point at which we have to deal with *Big Data* because the conventional BI data architecture is no longer sufficient. It has proven to be a good idea to use the integrated concept of the Data Lake architecture. This combines the classic structured BI architecture of the "Single Source of Truth" with the requirements of flexible analysis of unstructured data with advanced methods. From the staging area, the mass data are first enriched beyond their raw state and prepared for analysis and evaluation. The prepared data can now be used in so-called laboratories or sandboxes. This refers to environments in which analyses can be carried out safely, independent of daily business and the data stored in the data warehouse. These either help to answer specific management questions or are exploratory in nature. Data exploration refers to the iterative process with largely automated procedures for analyzing multidimensional data and determining the new usable information contained in them using intelligent methods from statistics, visualization, machine learning, knowledge discovery, and database technologies. These data analyses are not so much based on given hypotheses or defined questions, but rather the primary goal is to use methods of so-called data mining to support the viewing of data, the mental imagination, and the explanation of patterns or of hypotheses.

This means that an attempt is made to discover previously unknown connections. Here too, the data used is maintained and stored for possible multiple use over time. However, never with the intention of generating reports directly from this data, or to allow departments or management to access it directly. Work in the Advanced Analytics area should be reserved for specialists.

The results and findings of the analyses are used in two directions. On the one hand, they can be used for queries via the so-called serving layer. On the other hand, the results are used to transfer them to the core of the data warehouse for possible use via the data marts. In doing so, the data should of course follow the "classic" path via the staging and cleansing areas. The next step is the monitoring of real-time scenarios. Here a permanent data stream flows past corresponding receptors, which scan it for abnormalities and certain patterns. At a heating manufacturer, for example, Advanced Analytics analyses had shown that at certain combinations of sensor values, heaters tend to fail or consume disproportionately large amounts of fuel, because many of the systems are already connected to the manufacturer via the Internet. The manufacturer receives several thousand data records per second (!) from each heater. These are analyzed in real time in data streaming and the results are made available via the serving layer. If certain parameters are available, the manufacturer contacts the customer for preventive maintenance. The storage of the data from the *Big Data* is usually differentiated. Batch or near-time data are often stored permanently, for example to generate new analyses with modified methods and extended data sets or to map temporal progressions. Real-time data from streaming is usually stored for a limited period of time or only stored permanently in the form of regular "random samples". It is important in this concept that there is no redundant data storage. The "Single Source of Truth" also applies in the Data Lake. It is important to recognize that *Big Data* does not replace BI. Rather, it is necessary to do both the one and the other now and in the future.

Important competencies for controllers in the competence field of Business Intelligence are:

- Knowledge of the importance and the key points of a BI and *Big Data* strategy
- Knowledge of the basics and practical application of BI
- Overview of the possibilities of using BI for business planning
- Mastery of simple analysis methods and tools

Advanced Analytics

Advanced Analytics include advanced statistical methods and machine learning algorithms to extract information from unstructured and polystructured data.

Statistical methods are divided into two categories. Explorative data analysis comprises tools that are suitable for quickly gaining a basic understanding of a data set. The choice of the appropriate modeling technique depends on the knowledge gained so far. The tools used are contingency tables (comparing the frequency of values of different variables), scatter plots (as a standard tool for identifying the type of relationship between variables), histograms (empirical distribution of the values of variables as bar charts) and plots (violin/boxplot) as well as cluster analyses (division of data sets into similar groups).

With the help of inductive statistics, on the other hand, statements about the population can be made from the properties of a sample. For this purpose, random variables are generated and the probability of their occurrence is determined using probability functions. Important distributions are:

- Binomial distribution (coin experiment)
- Normal distribution (Gaussian bell curve)
- Poisson distribution (radioactive decay)
- Student's t-distribution
- Exponential distribution

By Machine Learning (ML) we understand the artificial generation of knowledge based on experience (Geron, 2019). All currently used methods have been known for decades. However, only the computing capacity and speed available today makes it possible to use them accordingly. Artificial intelligence (AI) learns on the basis of examples and can generalize, i.e. make generalizations after the learning phase is complete. ML is divided into four categories. Supervised Learning (SL) refers to the supervised learning of an AI. In the data set used, the result to be predicted (label) is known. Thus, the model can be trained and validated in a targeted manner based on historical data. The quality of the model is directly measurable by the deviation of the predicted value from the label. In contrast, with unsupervised learning (UL), the result to be forecast is not known in the data set. The model thus searches for common characteristics (clustering) or for anomalies or outliers in the data set. The quality of the model cannot be measured directly, but must be interpreted. Here the technical background plays a very important role. Controllers could use this method to detect, for example, anomalies in the fuel consumption of a vehicle fleet, catering expenses, or similar. Tax authorities already use this method. Even more sensible is certainly its use in connection with buyer behavior and the corresponding use in sales. With Reinforced Learning (RL), a model (also called an agent) learns an optimal strategy through targeted "rewards". Various objectives are pursued in the process. An agent is rewarded by solving tasks and receives more or less points depending on the quality of the solution. The agent should achieve as many points as possible by default.

- Search: The model looks for the shortest path to the solution
- Planning: The model looks for the most cost-effective path to the solution
- Game theory: An agent searches for the optimal moves considering an enemy agent

Examples include navigation systems or AlphaGo, the first program to beat a professional Go player under tournament conditions. AlphaGo used algorithms from both RL and SL. For Deep Learning (DL) so-called neural networks are used. These imitate the structures and processes of the human brain. They are able to learn complex interrelationships on their own, without explicit knowledge about the problem to be solved. Their application lies in image/pattern recognition, Natural Language Processing (NLP) and the like.

Important competencies for controllers in the competence field of Advanced Analytics are:

- Knowledge of methods and procedures from the field of Advanced Analytics
- Ability to perform advanced analyses of low complexity
- At least background knowledge of complex procedures in order to be able to communicate with data scientists in a goal-oriented manner

4. Define the *reporting* process

At which points in time reports are created in the respective company, whether there is a dedicated workflow, and whether there is a high or low degree of automation or not, is rather to be considered individually. A clear procedure with the highest possible technological support is certainly desirable here. This also includes clarifying who has which tasks and responsibilities at what point in time and how these are represented in the organization. At this point, in connection with the anchoring of the BI/*Big Data* strategy, the conflict between Self-Service BI (SSBI) vs. Governance, which arises again and again in the organization, should be emphasized. A useful definition of SSBI can be found on whatis.com: "Self-service Business Intelligence is an approach to data analytics that enables business users to access and work with corporate data even if they don't have a background in BI, statistical analysis or data mining. SSBI tools allow users to filter, sort, analyze and visualize data without involving an organization's BI and IT teams." IT has previously created the corresponding prerequisites for this by providing the data warehouse, the BI system, the *reporting* tools, and the self-service queries. However, the answer to the question of who exactly is meant by "user" is open to interpretation. We asked our participants in the CA controller akademie seminars about this. Seventy-three percent of the interviewed controllers saw themselves as users of SSBI, and 27% rather as the managers. The situation was very different with the managers surveyed. Some 92% of the managers saw themselves as users of SSBI, and only 8% as controllers. Exactly at this point there is a great danger for the actual idea of BI. If managers carry out analyses independently and completely freely, creating "auxiliary calculations" in Excel and other tools, the attempt to achieve a "single source of truth" very quickly fails. However, if every single change to a query has to go through the bottleneck of the IT department, the business department is not able to react quickly enough to the permanently changing conditions of the environment. The business department wants quick answers to questions on a user-friendly interface, even without IT. IT wants consistent data (models) in a secure environment. Both requirements are important and right. Clear data governance helps to bring both requirements under one roof. Data governance is a kind of house rules that define rules, authorizations, and processes for handling data in the company. The following governance approach has proven its worth in ensuring a "Single Source of Truth".

Authorization of the management:

- "Fetch" all available standard reports from the data mart
- Flexible analyses via intuitive filter functions of dashboards
- Analysis of existing drill-downs and report links

Authorization of the controllers:

- Customization of existing reports, analyses, and dashboards
- Creation of new reports and analyses (ad hoc)

Obligation of the controllers:

- Integration of the adapted and new reports and, if necessary, adapted data models into the data warehouse together with IT (or corresponding roles) after they have been validated
- Information to IT when reports and data models are out of date

This approach ensures that there is either a single central data warehouse, or possible decentralized "offshoots" in areas and departments, but with a common data structure, common data models, and, above all, common definitions. This should be clearly regulated in a data governance or business glossary. Of course, other "dosage forms" of reports are not obsolete for this reason. These can still be presented to a larger plenum in a meeting, explained individually to individual decision makers during a "home visit", or sent out. With constantly changing decision bases, the creation of a " daily manager newspaper" is also conceivable. While the controllers carry out analyses in a variety of ways and use dynamic dashboards to look for levers to achieve their goals, the decision makers receive the information relevant to them every morning without exception, with clear descriptions of the situation, analysis results, and recommendations for action. That would be service instead of self-service for the management! In this way the controller can become a real business partner. For this to work, the controller must have access to data and analysis tools and have internalized the business model and the objectives of the company. Last but not least, the ability to communicate the insights gained in a recipient-oriented manner is also required.

5. Communicate insights to the management

Management Information Design is – for all its fascination with calculations and technology –the last important step in building the bridge from the collection and analysis of (mass) data to management decisions ("From *Big Data* to Executive Decisions"). It is crucial to prepare the information obtained in a way that is understandable for management. This is where the IBCS® (International Business Communication Standards) SUCCESS rules for recipient-oriented preparation of management reports have become established in recent years (Hichert and Faisst, 2017). They ensure that reports convey a clear message and that the content is embedded in a story line in a meaningful way. Important aspects are to use a uniform notation, to increase the information density, to ensure visual integrity, and to use only the most suitable visualization. Excel can be used for the visualization within the framework of a frontend in the same way as corresponding tools from BI providers. In the context of the dynamics described above, the unified notation concept of IBCS is crucial, making it easier to understand rapidly changing reports and dashboards through pattern recognition. The same things (always) look the same. Different things look different (Hichert and Faisst, 2019: 10).

Important competencies for controllers in the competence field of Management Information Design are:

- Knowledge of the IBCS® SUCCESS rules
- Mastery of tools for creating reports in front ends based on the IBCS® SUCCESS rules
- Ability to structure storylines professionally and implement them in presentations

6. Data literacy as a key competence of all employees

The ability to deal with data in a planned manner and to be able to consciously use and question it in the respective context is called data literacy, which could be translated as data competence. Even if the operational handling and the analysis and interpretation of data is largely the responsibility of controllers and data scientists, all those responsible for managing the company should at least understand the information on which they base their business decisions. In uncertain times, many companies hope that the increased use of data and the application of advanced analytics will lead to improved decisions and valid predictions. Decision makers in management are often unaware that pure data collection and the use of algorithms are far from sufficient for this. Man and machine must "work" closely together. This is because the information that is ultimately used to make decisions depends on a variety of assumptions. These should be influenced by management or at least be known to them.

In addition, it is recommended that a certain level of data literacy be developed for all employees of the company. The target state could be described as follows: All employees have a basic idea of which questions in the company should be answered by data analysis. For this purpose, for example, a company-specific concretization of the above-described content "Obtaining and using data" is helpful. In addition, it should be ensured that the employees are familiar with the company's objectives and the underlying control model. On this basis, they regularly question whether the data they use in day-to-day business could have a meaning in the overall context.

They also request data from which they expect to be supported in their work. A central unit coordinates this information and ensures that data are used sensibly and effectively within the company.

In the future, data literacy will have a similar meaning to being able to read and write. However, it will be some time before it is introduced into school and university education. For this reason, companies would do well to build up this competence as broadly as possible today.

References

BENNETT, N. & LEMOINE, G. J. (2014) "What VUCA Really Means for You", Harvard Business Review, January 2014.

DOERR, J. (2018) "Measure what matters", Portfolio (24. April 2018).

GERON, A. (2019) "Hands-On Machine Learning", O'Reilly UK Ltd. (11. October 2019).

GORELIK, A. (2019) "The Enterprise *Big Data* Lake", O'Reilly Media (21. March 2019).

HICHERT, R. and FAISST, J. (2017) "International Business Communication Standards", IBCS Media (11. July 2020) Definitions, pp. V-VII.

HICHERT, R. & FAISST, J. (2019) "Solid, Outlined, Hatched", IBCS Media (31. October 2019), pp. 10-11.

WOHLAND, G. & WIEMEYER, M. (2012) "*Denkwerkzeuge der Höchstleister*", UniBuch Verlag (10. October 2012).

WOHLAND, G, (2020) "www.dynamikrobust.com".