MANAGEMENT PERSPECTIVES OF DATA-DRIVEN, ECOSYSTEM-BASED BUSINESS TRANSFORMATION

A Master's Thesis submitted for the degree of
"Master of Science"

supervised by
Dr. Larry Stapleton

Alin Kalam

01400355

Vienna, 01.04.2019

Affidavit

I, **ALIN KALAM**, hereby declare

1. that I am the sole author of the present Master's Thesis, "MANAGEMENT PERSPECTIVES OF DATA-DRIVEN, ECOSYSTEM-BASED BUSINESS TRANSFORMATION", 93 pages, bound, and that I have not used any source or tool other than those referenced or any other illicit aid or tool, and
2. that I have not prior to this date submitted the topic of this Master's Thesis or parts of it in any form for assessment as an examination paper, either in Austria or abroad.

Vienna, 01.04.2019

_____

Signature

ACKNOWLEDGEMENTS

ABSTRACT

From the first wave of mechanical production to current day, throughout modern history we have been witness of several industrial revolutions. As if it was not enough also ground-breaking political and socio-economic transformations related to those modernizations reshaped our world several times and for sure there are yet many transformations to come! In times globalization and emerging developing markets, organizations are facing numerous challenges. According to a Bitkom survey, 30% of businesses in Central Europe say they have trouble coping with digitization and the main reason behind this circumstance is their overall approach to data-driven innovations.

Since the amount of available data increases steadily, both private companies and public organizations are concerned with managing and handling data, which in many ways is the key to developing innovative business models across all sectors and technology fields. As a leading IT & Management executive, I have explored in last nine years digitalization as nothing new but much more an on-going process, which started decades ago with the mass automation of modern industry. For a long time, the extent of digitization has been therefore underestimated as a simple buzzword. However, it is clear that the digital transformation is more than only further automation, since it covers e.g. concept of user experience, emotion based automation and above all artificial intelligence supported through modern data-driven capabilities as machine learning etc. It even combines data-driven process optimization with revolutionary producing technologies as 3D printing, which is also known as "Industry 4.0". Having mentioned the current progressions, digitization will reach its peak very soon and every industry should be concerned about it. On the one hand digitization will help creating new business models and emphasize new market players with competitive advantage against legacy companies and on the other hand it is a door opener to process optimization, cost reducing and further automation for nearly every field of economy.

For instance unlikely many traditional companies', digital players such as Airbnb, Uber, Spotify, Netflix and co. put not only customer benefit, profitability- etc. but also modern data processing capabilities in focus. Mainly because only those capabilities deliver important insights and possibilities of predictions for decision-making processes, building new services and addressing them to customers. A dada-driven company from bottom to top can process data effectively to create valuable information & insights faster than the competitors and seen from this perspective most of digitization is data-driven.

This thesis will scientifically confirm many aspects & hypothesis of data driven-digitization. It will handle the topic of "data-driven digitization" directly connected to management & ethical aspects to provide detailed insights about synergy effects between those fields.

RESEARCH QUESTION

It is likely that Management Boards of multinational companies from IT/Telecommunication, public- & finance sector do nearly the same mistakes repeatedly when dealing and investing in innovative technologies to master the challenges related to the data-driven future. By leveraging my experiences and best practise, I have figured out mainly three areas of concern to prevent those and take regard of recent but yet widely unexplored concepts & technologies mainly form a management or executive perspective:

> Research Area 1: Contrasts between data-driven approaches & ongoing digitization/automation
> Research Area 2: Impacts of data-driven management, strategy & organization
> Research Area 3: Ethical aspects: opportunities & challenges

The aim of this research work is to analyse the current market situation of large companies in Europe. Therefore I have chosen mainly the European finance - and aviation industry to find answers for the research question:

"What are status quo, upcoming challenges, risks, opportunities & added value of data-driven and ecosystem based approaches and how do they correlate with ethical standpoints?"

In order to find suitable holistic answers with a broader spectrum to this research question, a number of sub-questions were formulated:

> What are the major synergies between data-driven digitization & ethical standings connected with strategic & decision making aspects?
> Market analysis with quantitative and qualitative research on best practice & real-life projects
> Which role play concept and implementation of e.g. "Big Data", "Data Lakes", "Data Ecosystem" within related organizations?
> What are the major impacts of data-driven digitization in the society and does it correlate with the situation in multinational companies?
> What are the key success factors effective ways to digitize the future
> How does decision making work today and what is the future?
> Does moral/ethical standpoints block innovations and data-driven development of today´s business?

> What is the preferred solution approach with scientific but also practical backing of this thesis? (of points mentioned above)

RESEARCH HYPOTHESIS

There are many challenges and opportunities of data-driven approaches existing today that are inevitable. Those may differ from the specific type of the organization in the first place but however the over-all challenges & impacts do have a major overlap and are nearly the same regardless the specific field of business. There is also a major synergy effect between private- & public sector in terms of changing company-, customer- etc. culture to improve da-to-day business and even raise effectivity in such organizations.
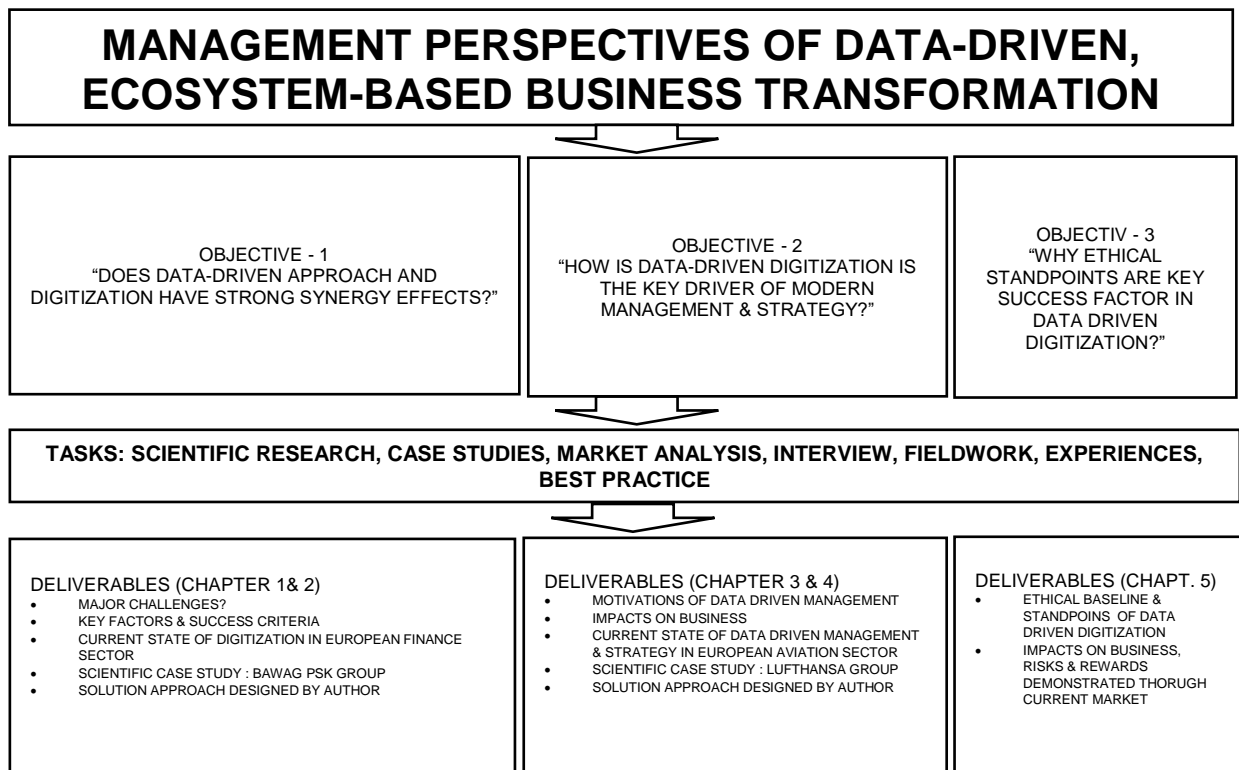
A particular focus is being put on the question how far companies are aware of these common but yet very complex innovations directly connected to ethical/moral & legal issues. In order to answer the research questions three major hypothesis are to be addressed in this thesis. Each research area will have consecutive introduction, market research, challenge statements, major impacts and further scientific insights followed by specific solution approaches produced through research work of this thesis incl. data from two multinational European concerns from banking and aviation field.

> Hypothesis 1 (Research Area 1): Data-driven approach is the core element of digital transformation with immense impacts on economical but also sociological fields:
> - o Solution approach : Implementation of data management & governance in an Austrian bank

> Hypothesis 2 (Research Area 2) : A data-driven management e.g. enabled through Business Intelligence and Data Science adds additional strength to business and fastens the digitization process of big companies:
> - o Solution approach: An example of an interactive dashboard built with data from an European airline

> Hypothesis 3(Research Area 3): Ethical Aspects of data-driven transformation when neglected build serious threat to existing business and can be destructive for future development

RESEARCH METHODOLOGY

In order to satisfy the objectives of the thesis, a qualitative research work was held, which delivers findings and conclusions that will produce results. The chosen approach is built on three major pillars, which are:

- Scientific research work & market analysis
- Interview with industry experts from banking, aviation and etc. sectors
- Personal experiences, domain knowledge and best practise

MANAGEMENT PERSPECTIVES OF DATA-DRIVEN, ECOSYSTEM-BASED BUSINESS TRANSFORMATION

OBJECTIVE - 1
"DOES DATA-DRIVEN APPROACH AND DIGITIZATION HAVE STRONG SYNERGY EFFECTS?"

OBJECTIVE - 2
"HOW IS DATA-DRIVEN DIGITIZATION IS THE KEY DRIVER OF MODERN MANAGEMENT & STRATEGY?"

OBJECTIV - 3
"WHY ETHICAL STANDPOINTS ARE KEY SUCCESS FACTOR IN DATA DRIVEN DIGITIZATION?"

TASKS: SCIENTIFIC RESEARCH, CASE STUDIES, MARKET ANALYSIS, INTERVIEW, FIELDWORK, EXPERIENCES, BEST PRACTICE

DELIVERABLES (CHAPTER 1& 2)
- MAJOR CHALLENGES?
- KEY FACTORS & SUCCESS CRITERIA
- CURRENT STATE OF DIGITIZATION IN EUROPEAN FINANCE SECTOR
- SCIENTIFIC CASE STUDY : BAWAG PSK GROUP
- SOLUTION APPROACH DESIGNED BY AUTHOR

DELIVERABLES (CHAPTER 3 & 4)
- MOTIVATIONS OF DATA DRIVEN MANAGEMENT
- IMPACTS ON BUSINESS
- CURRENT STATE OF DATA DRIVEN MANAGEMENT & STRATEGY IN EUROPEAN AVIATION SECTOR
- SCIENTIFIC CASE STUDY : LUFTHANSA GROUP
- SOLUTION APPROACH DESIGNED BY AUTHOR

DELIVERABLES (CHAPT. 5)
- ETHICAL BASELINE & STANDPOINS OF DATA DRIVEN DIGITIZATION
- IMPACTS ON BUSINESS, RISKS & REWARDS DEMONSTRATED THORUGH CURRENT MARKET

TABLE OF CONTENTS

CHAPTER 1.0

1.1 RESEARCH AREA 1 - Contrasts between data-driven approaches & ongoing digitization/automation

Hypothesis: **Data-driven approach is the core element of digital transformation with immense impacts on economical but also sociological fields.**

Despite many progressions of last years, the western economy has been standing in front of a vast amount of challenges because the true digital revolution coupled e.g. with AI, automated predictive analytics or unsupervised Machine Learning & etc. is yet to come.

Detailed market analysis, research work and also years of professional experience in telecommunication, banking and tech-driven sectors had shown me that yet very fundamental steps must be undertaken to even start with further optimization and automation of human based business areas. Most of the major multinational companies often do not even understand fully the benefit of digitalization to their technical but also business ecosystem, which makes them vulnerable in many ways.

5 KEY FACTORS OF CHANGE IN THE DIGITALIZATION ERA

Macro-economic conditions	Changing customer behavior and expectations	New competitive forces	Ongoing Digitalization & Automation	Regulations & other factors

OUR POSSIBLE RESPONSES

Optimize & create new business by offering more **DIGITALIZATION** based on **SINGLE POINT OF TRUTH** for data

Generate **ROI** by using BI, Data Science and related technologies

FLEXIBLE & UNIFIED data ecosystem to enable innovations

DEMOCRATIZE data throughout business departments to become a data driven company

Picture 1 - © Alin Kalam

1.1.1 Synergies of digitization & data-driven

Our world has never seen before such data & information driven society in its history than right now. As it was predicted many times, ninety percent of the data in the world today has been created in the last couple of years alone. According to a recent research by Domo [1] currently our output of data is roughly 2.5 quintillion bytes a day. As the world steadily becomes more connected with an ever-increasing number of electronic devices accessing the Internet that is only set to grow. By providing internet, access to nearly every parts around the globe we are not only achieving democratization of information but also impacting the economic landscape.

Seen from this perspective **"the core of digitization is data-driven"** because nowadays most of business applications, services or other touch points between customer & provider generate vast amount of data that tells us a lot about e.g. customer behaviour, propensity to buy next best product or even customer satisfaction.

After years of development, it is now possible to build capable technical platforms and processing methods of data/information ecosystems e.g. Big Data technologies that are mentioned in further chapters below. Above all the handling of data e.g. modern analytics, predictive maintenance, Machine Learning etc. solutions and services built on those are shifting traditional & offline customer segments towards "digital native" beyond the natural growth of this segment. Current market situation of streaming providers such as Netflix & Spotify show how fast traditionalist key players are dethroned without having even enough time to react. In fact, the survivors of this shifting phase will be those who understood "to do business as a tech company" and undertook the right steps to push forward and reshape their business processes but also product portfolio constantly. As example the company Netflix could investigated initially to show how data-driven approaches drive modern-day innovation sin companies.

1.1.2 Status quo – Netflix & data-driven AI revolution

Netflix is a technology giant which spending is mainly on Artificial Intelligence to make content spending decisions. It improves its algorithms and uses them to reduce the human intervention in programming decisions. Netflix operates its own CDN (Called open Connect). [2] This helps Netflix save bandwidth cost and enables access from the same region throughout multiple consumer groups. Netflix has also a backup of the content on cloud storage which can be recovered just within in hours if data mishandlings happen.
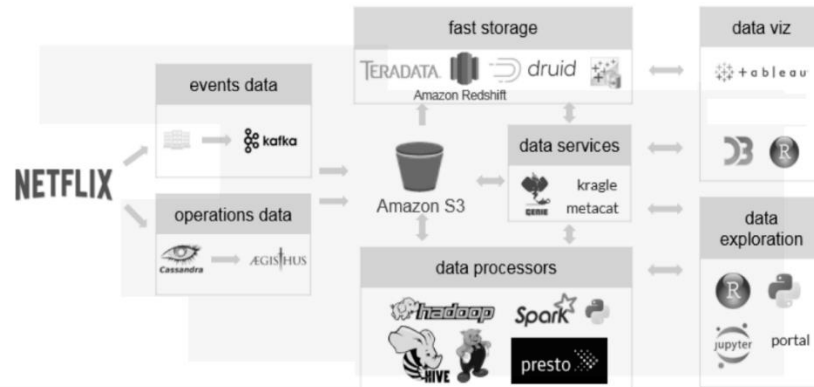
Picture 2 - © Tableau.com

The analytics platform mentioned illustrated above has made Netflix leader among its competitors. In many terms Netflix to date the only content generation company that mostly prefer data and analytics at this scale. The major impacts of its data-driven approach are:

> The success rate of Netflix shows are beyond 80%

> Netflix provides a customer experienced driven personalized experience

> According, to many surveys it shows that around 75% of it´s viewing is driven by the artificially evolving recommendation algorithms

> Their use of data and prediction of the success rate of the shows has helped Netflix to reduce a lot of cost on promotional advertisements and campaigns

> Higher customer satisfaction

It is also worth mentioning that data-driven digitization is not only a matter of private companies. In case of India is in many ways also a major driver of modernization of society effecting lives of millions of people.

1.1.3 Detailed market analysis – Data-driven digital transformation of India

India is one of the countries where cash payment is traditionally rooted and this circumstance has been a recipe for persistent corruption throughout the whole country since decades. End of 2016 the government was forced to invalidate 500 and 1000 Rupee banknotes overnight to fight corruption and black money, which had accelerated the digitization in the country as a positive side effect. Research work of on this specific example from Dr. Ajay Pratap Yadav, Dr. Awadhesh Kumar Tiwari [3]demonstrates, that data-driven digital transformation is not a topic of revenue generating private companies only but also very relevant to development of societies and countries like India.

1.1.3.1 How India´s digital transformation was boosted

Beside many progressions & development successes that India has been lately experiencing, it is still a poor country [4] (according to "Global Finance" Rank 126 out of 189 countries) with a lot of cultural-, political-, economic-, -environment- but also sometimes a population problem. Even though it is performing very well and working hard to decrease poverty, down to the day a huge amount of citizens do not have access to electricity, fresh drinking water or even basic infrastructure.

Millions of hand-to-mouth citizens from poorer segments of the society are not used to have a bank account or cannot correlate it with their everyday needs. Nevertheless the "note ban" of 2016 has been forcing millions to get bank accounts. According to government guidelines it was possible to exchange up to 4,000 rupees per day until 24th November, (anything over this would have been subject to tax laws). So citizens could deposit their cash money to their accounts and alone for that fact a nationwide rollout of bank accounts was needed, which is a step forward for itself because all of a sudden several doors open up at the same time to move faster forward towards "Digital India".

Picture 3 - © PAYTM

It was nothing but a great achievement to provide digital payment services to overcome the absence of infrastructure in rural low developed areas, where is not possible to build electricity connections or baking facilities for the poorest. Apparently these segment of the society below poverty line is even today one of the biggest sufferer of fraud notes as they already have very
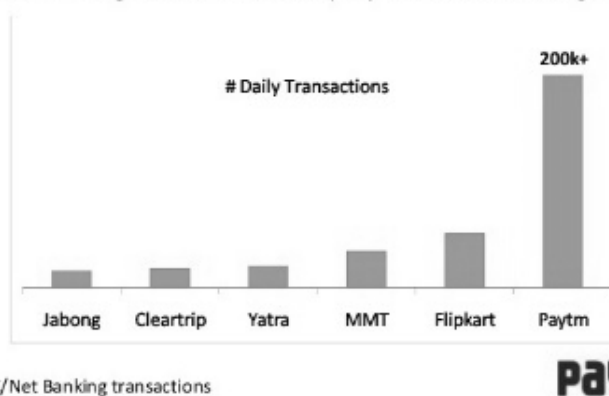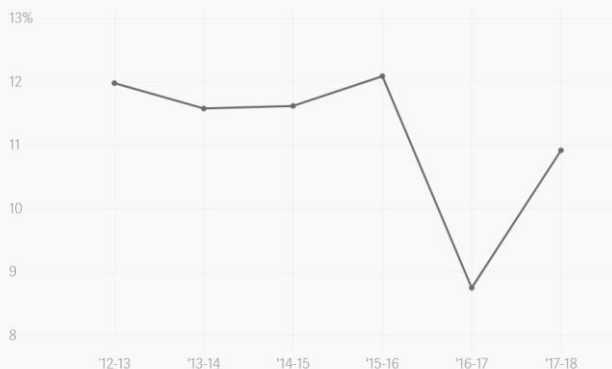
Picture 4 - © Bloomberg

less and earn day-to-day for survival. So enabling this segment of the society towards free digital banking services supports the evolving super power in Asia. Provider as "PAYTM – Pay through Mobile" (Picture 3) is such an example how even a country with huge population can reduce paper notes. Besides solving black money a few but yet important problems such as saving resources or providing flexibility to a huge amount of people can be seen as a milestone of digitization.

One of a few lessons of this initiative is that such projects need a way better forecast and preparation than India had. Knowing upfront about the huge turbulence in money market and potential bottlenecks of national bank´s "slow" money printing capacity would have helped a lot to reduce endless lines in front of ATMs or desperate suffering of millions poor people. But nevertheless unwillingly these circumstances took again millions to alternative digital payment methods like E-Banking or Mobile Payment.

1.1.3.2 Aadhaar - A Big Data project for a billion combined with financial services

[5][6]Aadhaar is a 12 digit unique-identity number issued to an Indian resident based on their biometric and demographic data and the biggest worldwide of its kind. It should help to create an urgent consistent fundament for all citizens and this is used as gateway to government facilities like housing, ration card, healthcare etc. but also digital payment platforms such as PAYTM.
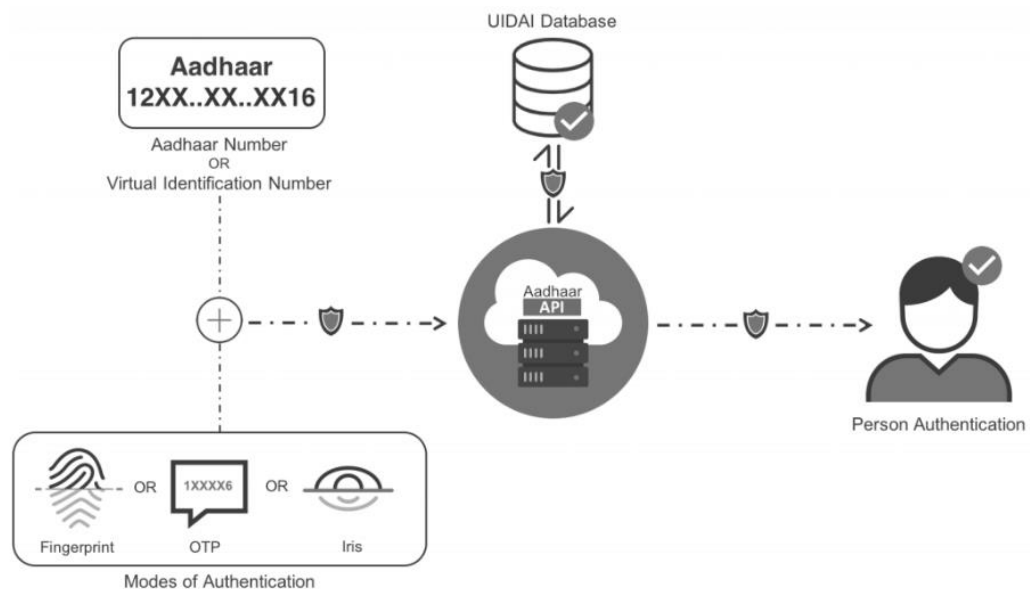
Picture 5 - © Aaadhaar website

At 1st look nothing very special at first sight but when we take regard of India's underdeveloped technical infrastructure, which makes wide-ranged or comprehensive national projects way difficult as it already is. A huge population of more than 1,252 Billion (status: 2013) spread out in very different

culturally/religious (caste system) diverse zones makes state-of-the-art technical implementations way complex than we are used to in the western hemisphere. Aaadhar is a truly big project developed for the population of India and it offered all of a sudden possibilities have a registration of birth etc. even for the poorest people [5] [6]. Beforehand they had been cut-off form the financial services and even government facility. With rising usage of digital payment services bridged to central data system Aadhaar, government and private companies have been able register every citizen uniquely and offer unprecedented services to this segment of society and beyond.

1.2 Summary

In case of India, it is demonstrated that a foresight to every data & digitization related developments is essential. Often this fluent connection between data-driven innovations and da-to-day business is missing. For most of the part, it is caused by lack of understanding the big picture regardless whether we take big companies or countries in account.

As a result, decision takers are of the opinion that simple investment in upcoming newer technologies is enough to jump over problems & challenges existing at the fundamental level which can be e.g. process-, data quality-, culture- but also effectivity driven. Instead of doing things right unfortunately, movement is often confused with progress. Many existing cases such as "Ebay & Skype " [7] or even "Nokia" showed us in previous years, that companies often act with gross negligence when it comes to understanding the core business and further development of it with a broader sight. In an ever-changing fast transforming world firstly companies, organizations have to break free from old schemata of implementing new approaches and doing business around those.

CHAPTER 2.0

2.1 Scientific investigation: Digitization of Central European finance sector

In order to substantiate the hypothesis, in following chapter the Central European finance sector incl. legacy banks will be investigated thoroughly. There are several mechanisms to rate and rank the grade of innovation within the finance sector. One of the well-known is [8] EFMA– Accenture Banking Innovation Awards, which is an initiative that aims to identify and award the most innovative projects in the retail banking sector at a global level, data-driven innovation will be the major challenge of global Finance sector for upcoming years. Mainly because nearly every innovation approach has major touch points with either data ecosystem - in terms of fast adopting and capable

architecture, data handling – in terms governance, security, access and even data processing – in terms of automated analytics such machine learning, data mining etc. The major innovation criteria are:

> Best new product or service
> Digital marketing
> Big Data, analytics and ai
> Customer experience & engagement
> New business ecosystem
> Workforce empowerment & behavior
> Wallets & payments
> Global innovator of the year (both challenger and established players)

Research works show that in comparison to other regions, Central European finance sector is years behind (Picture 6) in discovering and using opportunities through digitalization efficiently with the current backend IT setup. At the same time current market developments prove, that "traditionalists" have to deal more and more with FinTech-driven uncertain developments but also with constant "threats" outgoing from multinational companies (Apple, Amazon etc.) getting into banking & financing as we knew it so far. According to the research company Statista, 80 % of all "point of sale" transactions in Germany in 2016 were made with cash, compared with 68 percent in France and just 46 percent in the Netherlands. Countries like Austria are even beyond 80% mark, which has mainly a cultural reason. The usage rate of online banking in East European region is even way lower than e.g. Western Europe or even Baltics due to economic catch up process (e.g. integration into EU etc.).

Source: Finalta, Digital and Multichannel Benchmarking Study, 2016
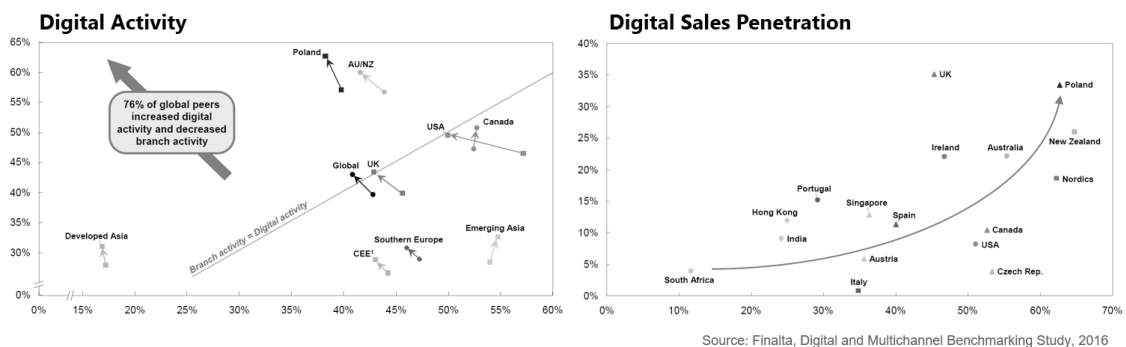
Picture 6 - © BAWAG, FINALTA Study

As if it was not enough revolutionary concepts and technologies such as Blockchain can be seen as another nail in the coffin. In other words without having an overall concept or roadmap towards further progressions of technical capabilities paired with business transformation in digitization domain, many traditional banks will vanish in future or forfeit most of their potential.

Austrian online usage	Reasons for not using mobile Banking	
	Too unsafe in general	36%
	Risk of economic damage caused by hackers	18%
	Data access	17%
	Don't need it	33%
	Too complicated	9%
	Existing offers are sufficient	6%
	Prefer personal contact	25%
	I do not know the offer of banking online	3%
	Other	18%

71% „Security concerns"

48% „No additional value add"

- Onliner and mobile banking users
- Onliner but not mobile Banking users
- Offliner

Potential of 26% of Austrian population that are online, but not using e/m Banking, to digitally engage

Source: AIM study, 2015

Picture 7 - © AIM study

2.1.2 Challenge: Big picture

Recent market trends prove once more that many banks have not fully understood the value of technical and business ecosystem and they will most probably won´t be able to keep up the speed of current wave of transformation. A look into previous yearly figures might show that bigger banks have an over-all functioning and ROI generating solid fundament of business built on backend legacy IT systems such as Cobol-based mainframes or Data Warehouse & batch orientated systems. Nevertheless this appearance alone is deceptive as soon detailed analysis are done, which show that those systems are generating lot of costs but also duplicates and inefficiencies from all corners. On top of that many have massive lacks in Data Governance but also quality in both master and metadata, which hinders pushing progressions forward towards prescriptive analytics e.g. to boost the omni-channel approach or even the digital customer base.

Implementations, any kind of adaptations in mainframe or regulatory or security projects generate lot of effort (as the case of American banks will be mentioned in chapters below), which should have been maybe invested into digital roadmap or transformation. Many direct banks as mBANK from Poland that are built on a solid fundament of data-driven ecosystem, show how efficient and effective a pragmatic yet focused approach can become.

2.1.3 Challenge: Wrong assumptions

According to the industry review 2016 of Cloudera– a major Big Data developer, in 2018 nearly 85% of data migration projects will fail either due to

missing use cases or lack of long-term strategy. So whenever bigger legacy banks are caring about reshaping & designing their omni-channel approach or adapt the customer journey/experience throughout their business, they experience a rude awakening lately by losing customers and market share. The origin of this shaking disaster lies many times in the common assumption such as:

- digitization means cost reduction without taking care of e.g. customer experience & service orientation
- adding new tools to the ecosystem will solve major problems even basic issues e.g. process-related problems are not addressed
- inconsistency in extending the data ecosystem without having detailed concept about near future technologies on a roadmap
- data-driven innovation is innovation but not the key factor for digitization

Yes, data-driven Digitalization creates synergy effects between automation and cost saving but this fact must not be the major goal. Only by closing/ruthless optimizing of retail segment because mobile apps are launched and more customer data can be collected to boost online selling, it is hard to make room for better customer experience. The winners of Industry 4.0 & Digitalization era will be those who have understood the potential of combining online and offline with a proper strategy but also business cases.

2.1.4 Secret to success: Data handling

Central European Banks & other finance companies nowadays are simply forced to evolve data handling out of a "beyond the horizon" approach without knowing what comes next.   It has become inevitable to build a 360° data model for the logical and physical view of existing qualitative and quantitative data sets, which will be the core point for selling digital products, host historical data, facts and figures, making sure data quality, boost of existing business & etc. Such systems should be created to address given requirements combined with proper patterns and best practices of market. One of the key aspects of this area is to structure and simplify the whole process of data acquisition.

As if that would not be enough, there are constantly newer requirements as "real-time" data handling, which puts IT and business in front of huge challenges. Mainly because the vast majority of data ecosystems are based on **"batch data processing",** that is built by legacy systems and causes lots of latency in processing to deliver sophisticated real-time services in omni-channel (use case illustrated in picture 8)
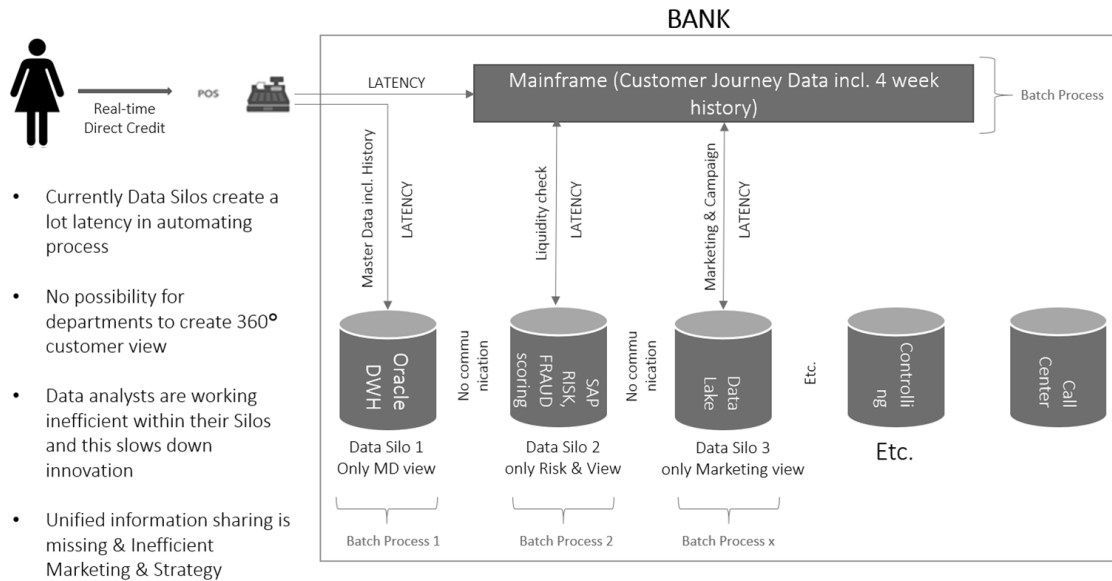
Picture 8 - © Alin Kalam

One particular answer to that latency issue is the so called **"streaming Data Processing"** (lambda & Kappa architecture). It has become a very important factor for so many business applications with customer touchpoints, which is literally the heart of all digitization & customer satisfaction.

The "streaming data processing" approach is originated in Big Data technology and remove inevitably latency from data systems (batch data processing as mentioned in picture above). It is used to query continuous data stream and detect conditions, quickly, within a small time period from the time of collecting/receiving the data. The detection time period varies from few milliseconds to minutes. For example, with stream processing it is possible build real-time responding services such as "fraud detection" or "propensity to buy" etc., which can deliver a major additional value for the customer. Streaming data processing is also called real-time analytics, streaming analytics, complex event processing etc.
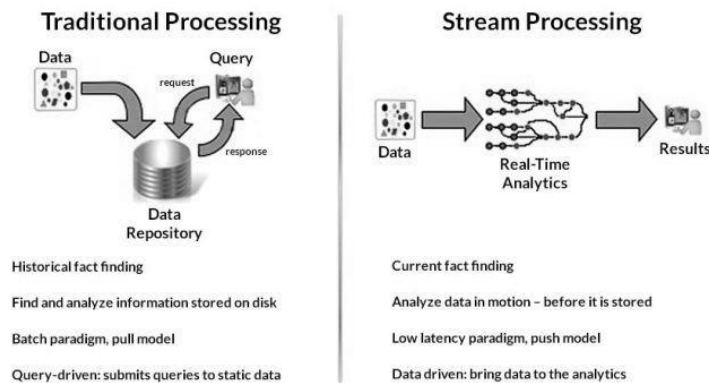
Picture 9 - © IBM

2.1.4.1 Financial impacts of Legacy

According to a Gartner industry paper (2018) nearly 70% of the yearly budget is being spent to keep legacy systems in operation and about 30% in investment in new technology infrastructure and new data-driven application development, which has been increasing since years. Recent data, industry journals and trade papers as The Future of Retail Banking Legacy IT Transformation by ORACLE [9] show this trend of increasing cost e.g. in case of other continents beyond Europe as well.

Deutsche Bank's information technology costs have risen for three straight years

Source: Deutsche Bank filings

Bloomberg

Picture 10 - © Bloomberg
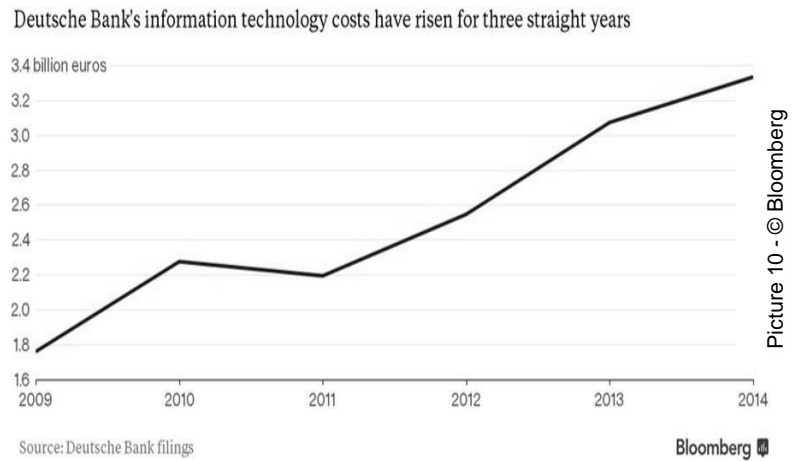
In the evolving process of becoming data-driven, bigger CEE Banks as Deutsche Bank, Raiffeisen International or Erste face enormous technology/IT migration challenges that are curtail to future developments. The rise of FinTechs, the emergence of new technologies as Big Data, block chain and faster networks. All of these can be a hind to the fact, that distributed computing is here to replace the centralized legacy computing. Particularly with the emergence of cloud and mobile computing, mainframes most probably will soon become an obsolete technology.

Sum of IT Spending

Picture 11 - ©CenterState
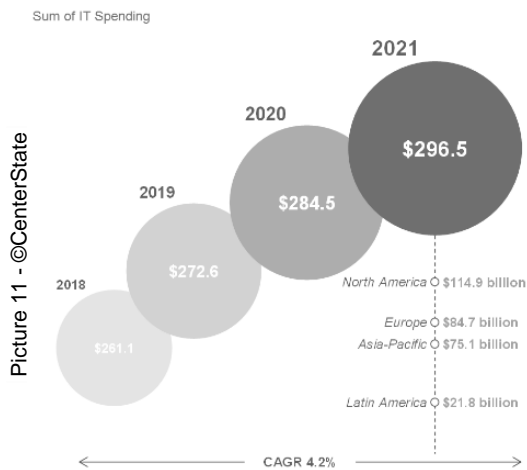
Out of complexity reasons of migration into newer technology, banks tend to keep their "batch data processing" based legacy systems and invest a lot of money to adapt them to newer market conditions. This fact beside many others is the major root cause of increasing IT expenses without having a true impact on the data-driven innovation & product development side. Closer look on the technology spending from another perspective, about 65% of the budget goes into just keeping the legacy. This amount goes to the basic infrastructure such as core applications, servers, security and data/customer driven business applications. These expenditures are usually not associated with revenue producing activities but are the foundation of the banking entity.

About 23% of the annual budget is targeted at existing business to grow income e.g. lending, cash management and other lines. That leaves only about 12% for new products, new processes and major business transformations such as automation, digitization or even building sophisticated customer experience to boost brand loyalty [11].
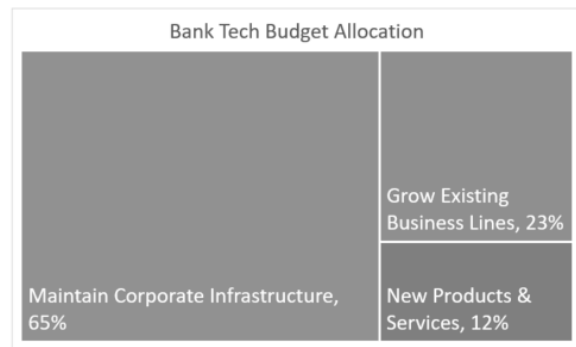
2.1.4.2 Why majority fails to evolve into "streaming data processing"

In a multi system environment maintaining both batch and streaming ingestion processes of multi-structured data is not a simple task. The technological challenges and hurdles in implementing such technologies are enormous. Even bigger companies like IMB, ORACLE, and MICROSOFT etc. are latecomers in this very field driven by open source driven innovations.

In addition to that there is a huge complexity in keeping legacy and creating new environment in parallel. Such an approach costs money and generates barely any revenue on a short term.

Many C-Level decision takers in finance sector miss out to think and invest on a long-term perspective even though technology is short-cycled than ever before in the history of mankind.

2.1.5 Data ecosystem: The key factor to success

Data goes nowadays far beyond the corporate boundaries. To date only a few handful legacy companies are able to use the full strength of data-driven business transformation. However current market trends show that companies have been consequently working to build data ecosystems to handle cross-border businesses transformation based on data. The ecosystem based data-driven mindset usually describes and develops the target architecture [12] [15] for those mid- to long-term strategic goals effecting business for many years. That's why beside data itself the ecosystem approach can be seen THE major success criteria on the journey to digitization. Ecosystems were originally referred to as IT environments.

It was mentioned in previous chapters, data-driven digitization is less about investments into infrastructure only but a lot about putting customer needs in center based on modern technologies and useful automations rather than only cost saving agenda. There is no such need of only investing in newer technology when the digital products do not offer added value to emotional

experience-, satisfaction etc. of customer through better services. In case of many technology giants as Microsoft or even Amazon, it shows how effective foresighted ecosystem approaches can be. According to the "Enterprise Cloud Strategy" paper, by end of 2017 Microsoft did not only change their business strategy towards cloud but also reshaped their whole business model & product portfolio dedicated to its cloud platform AZURE [27]. It clearly focuses on such ecosystem approaches by saying:

"With continuous optimization, you are always on the lookout for overprovisioned servers, for underutilized databases and servers, and for applications that are lightly or not at all used in off hours. For a large IT ecosystem in the cloud, this can save literally millions of dollars".

[28]According to the Annual report of 2018 Microsoft invested nearly 23.2 billion US dollar in newer cloud-based products, which has increased by 56% than previous year. At the same time Server products and cloud services revenue increased 21%, driven by AZURE revenue growth of 91%. This has been the major driver for Microsoft recent "catch up race" in keeping and growing new market shares. Also in case of Amazon it is more than proven that without an ecosystem based approach the company would have had a major gap in revenue (picture 13). It

Picture 13 © Geekwire (Amazon)
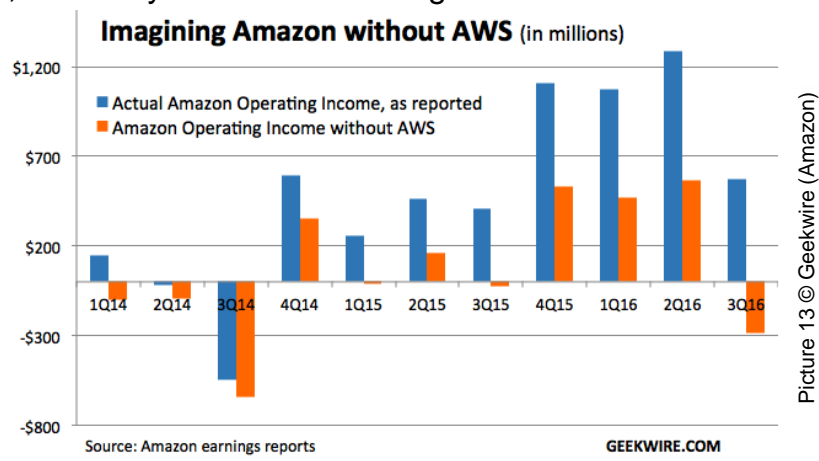
is worth to mention that, both companies has not only transformed themselves, but together with google as technology providers they can be seen as the leader of "ecosystem thinking" with their cloud-based product portfolio, which can be seen as a living example of ecosystem-based
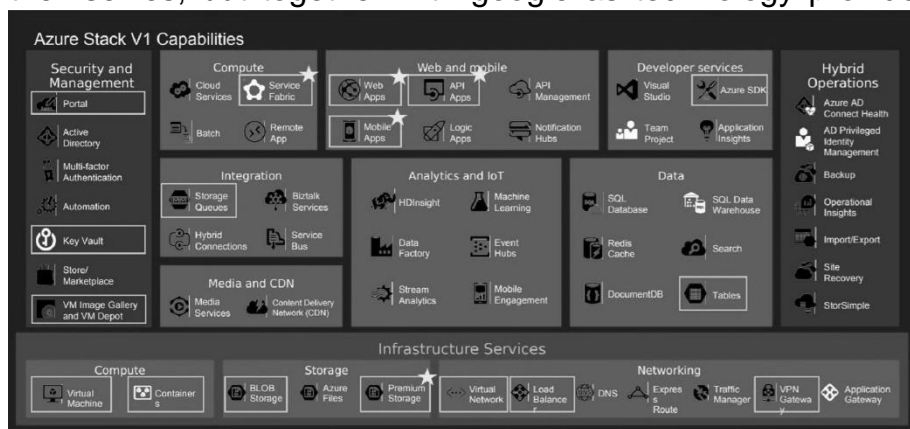
Picture 14 © Microsoft

digitization of their customers. In case of AZURE "Data Ecosystem Stack" (picture 14) it is well demonstrated how those companies drive ecosystem

orientated "software as a service" business model, where customer can choose form a vast amount of tools and setup out of an enormous services.

Thus historically seen Data Ecosystems had been designed to be relatively centralized and static. The birth of the web, cloud services and above all collection and usage of data has changed it radically. Now, data is captured and used throughout organizations and IT professionals
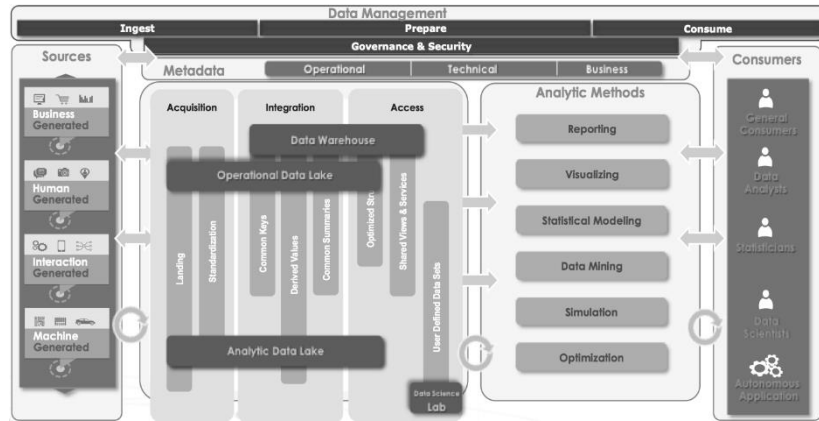
have less central control. The infrastructure they use to collect data must now constantly adapt and change. Hence, the term data ecosystem: They are data environments that are designed to evolve and there is no such standard or one solution for that, which is the reason why Every business has to create its own ecosystem, sometimes referred to as a reference model as in picture 15 and fills it with a patchwork of hardware and software to collect, store, analyze, and act upon the data. The most functional and useful data ecosystems are built around a product analytics platform that ties the ecosystem together. Analytics platforms help teams integrate multiple legacy data sources as e.g. core banking systems, and provide modern data processing & machine learning tools to automate the process of conducting analysis.

The so called "target solution" of such ecosystem is therefore often the central data hub for analytics, operations and services. It is more than standard legacy platforms as today, because it covers needs to go further than data storage and access to its raw form, but needs to cover also integration with business applications and services. We can distinguish three layers of the architecture which build on each other [12] [13] [14]:

- Data Lake – main layer providing and organizing data from the source system. At this level data is delivered do the hub, has passed initial validation, has been curated and stored in almost raw form, which means that some standardisation may occur but no business logic changes or integration took place at this stage.

- Data Reservoir – at this level the "raw" data is provided by Data Lake
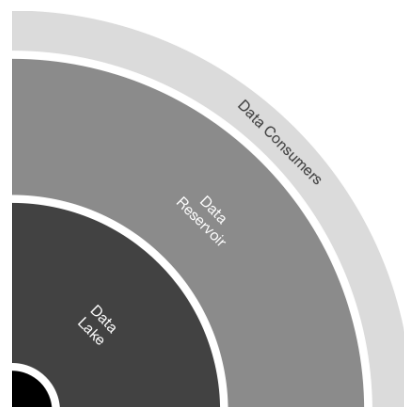
layer and can be now optimised and prepared for particular business needs. At this stage we make sure that necessary integration of the data sets is provided with proper metadata management, data registration, data lineage and access interfaces.

- Data Consumption – in this layer we have all consumers of our hub. From the data point of view our previous two layers have provided the full stack - the data is prepared and there is a need for space in which applications can integrate with the hub.

2.1.5.1 Data Lake

According to Gartner a Data Lake is a collection of storage instances of various data assets including the originating data sources. These assets are stored in a near-exact or even exact copy of the source format. The purpose of a data lake is to present an unrefined view of data to only the most highly skilled analysts, to help them explore their data refinement and analysis techniques independent of any of the system-of-record compromises that may exist in a traditional analytic data store (such as a data mart or data warehouse).

1. Source
2. Ingestion
3. Acquisition
4. Core
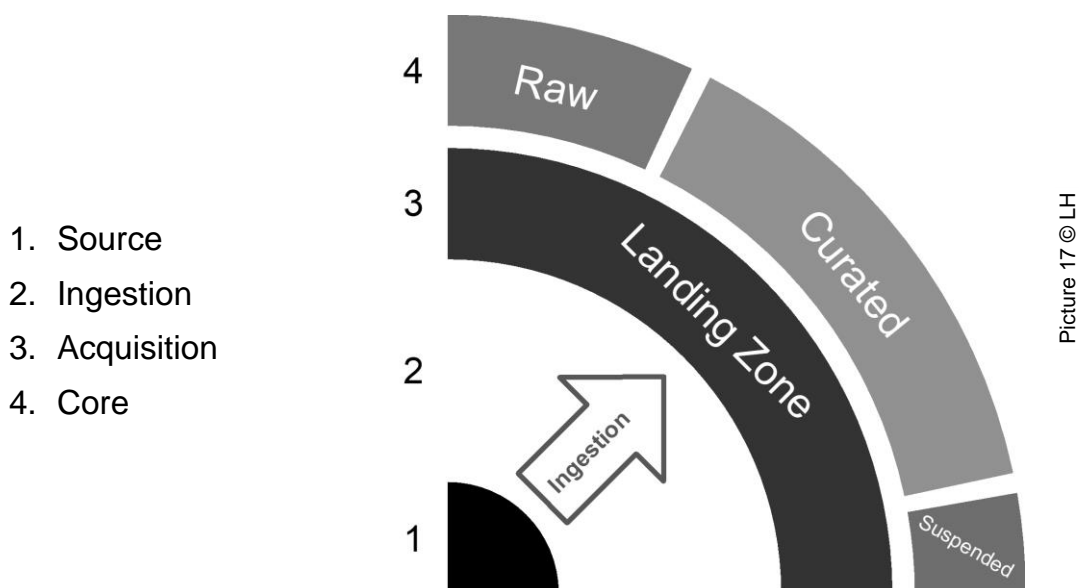
Ingestion is the process describing data movement to the Data Hub. It includes the proceeding of getting data from the source system, moving it onto the data lake in its raw form or as raw as technically possible as well as placing it in an initial location. The commonly is underestimated as connectivity and interface implementations are often not planned for appropriately. Assuring existing interfaces are known and technical connectivity between the systems is both possible and plausible (i.e. data is not unnecessarily transferred between too many unneeded network hops).

Acquisition is a transient area designed for intermediate validation and pre-processing. Here incoming data is technically and possibly logically validated,

assuring the sanity and quality of new data. Validations and curations of the data in this layer are very generic to assure it is still usable by a wide range of business areas. Running processes in this early stage of the data eases later business processing and allows for multiple lines of business to share their useful data initiation procedures.

Core is the permanent data storage layer for keeping both archival and operational data. Non business specific parts of data processing or any processing valuable to the entire business takes place here. Any incoming data is made available through the separate areas, curated, raw and suspended. Where curated is the finished data ready to have products build on top of further joining and investigations done with. Raw is the provision of the untouched incoming data, to allow for business specific applications to have full non abstracted access to the incoming data and be fully transparent about the proceedings done on the data this far. Suspended is an area for data considered invalid during the incoming data processes. Here businesses and data stewards have the chance to find issues with the data and think about solutions to clean up incoming streams, sometimes these can even lead to findings.

2.1.5.2 Data Reservoir

Atop the data lake, the meanings of a data reservoir become clear. The previously described now available ingested data is accessed in different technical means. Each method of access can be built on by lines of business. Each LOB can thus create its own area and prevent interrupting other ongoing projects. An LOB area can also share its success, finished new data products or newly consolidated data with other LOBs, easing their future and current efforts.

Projection

Applied

Access

Picture 18 © LH
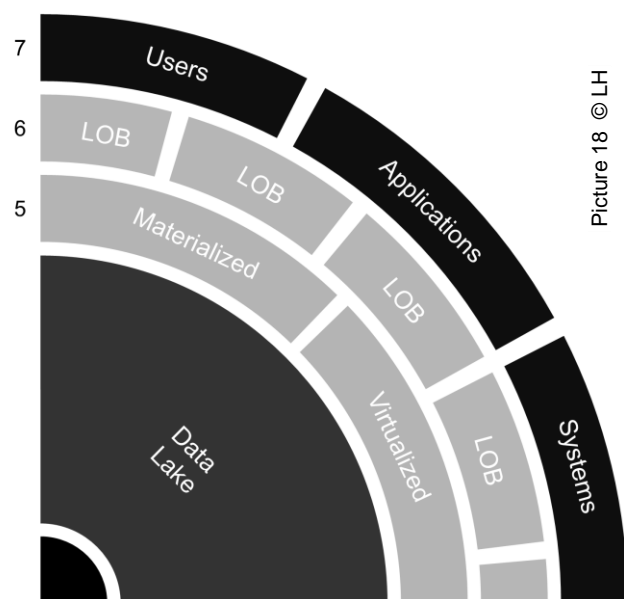
Projection is the layer that enables data for general consumption. This layer merely describes the mode of consumption or the matter in which data is accesses. A materialized consumption of access is the classic working with data, where both the access as well as the processing resides on the same locally resident data. By locally we do not mean local the accessing machine, but local to the distributed data lake. Materialized data has already gone through the entire process of getting data onto the data lake (as previously described) and is most suitable for operations within the data lake. This data is served within the infrastructural capabilities as the data lake and thus assures well performing processes.

Applied is the area describing data with applied business rules and logic. Here we find different line of business areas (LOBs). A LOB can be zoned per business area or per team or even per project, depending on the usage requirement of the LOB. An LOB can also be based on materialized as well as virtualized access, always considering performance requirements and the technical solutions behind the data. Within an LOB the interacting users enjoy the freedom of being abstracted from other LOBs, thus not intervening with other ongoing projects. Input and output of an LOB as well as containing data and processes of an LOB are entirely up to the LOB owner and the requiring projects. Tools used within the applied layer should however be within the stack chosen in the Solution Building Block section **Error! Reference source not found.** or considered under special request or as part of the Analytical Target Architecture (in a separate Document).

Access describes data objects, which are optimized for consumption. LOBs producing data artefacts, data products or simply newly available data objects have to provide them in a way consumable by data scientists or applications or the business itself. LOBs can also be of investigative nature, in which case access to their results may only be needed by the line of business itself. However we should strive to create reusable artefacts and results, to assure future improvements find effect in as many areas as possible.

2.1.5.3 Data Consumption

Finally we see a detailed view of all layer described previously. We also show the final step, the consumption layer, this is usually not only part of this target picture but works intertwined with the overall Data Architecture. The consumption includes Data Science efforts, here mentioned as Data Science lab, however these do not need to constrain themselves to a lab-like environment and can also be fully working production procedures being prepared for the later required production automation. Another part of consumption is classic BI Tools and advanced analytics. This stands for all

reporting capabilities and provision of data and insights to the business in a manner already established with them, the classic BI or interactive report. Such reports can both be provided by production processes as well as being self-serviced by the business departments. The next a largest player in consumption is the service layer consumption. This layer includes all automated interactions and interfaces provided to other e.g. partner ecosystem's applications

1. Source
2. Ingestion
3. Acquisition
4. Core
5. Projection
6. Applied
7. Access
8. Consumption
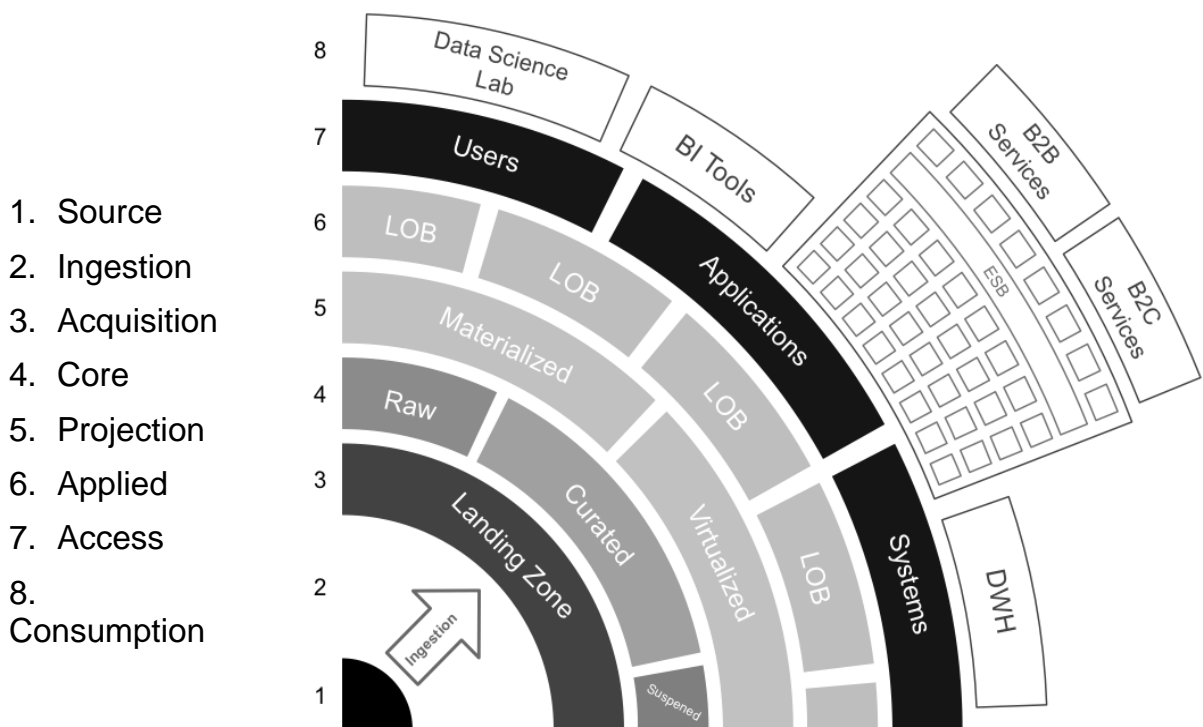
Picture 19 © LH

2.1.6 Summary

In a nutshell three major areas of a data ecosystem should be infrastructure, analytics & applications. The example mentioned above only might solve the issue of data storing and handling within a technical infrastructure. In a matter of fact a best practice data ecosystem should allow companies to store, process, analyze and visualize data adapted their business need, which is a very hard tings to achieve. Therefore the connection between the "business ecosystem" and "data ecosystem" can be seen as the holy grail of digital transformation out of complexity reasons of legacy technologies. On one hand companies must keep functional legacy systems to run daily business and develop new and on the other hand constant market and customer needs force them to invest in newer technologies to solve challenges of digitization era. It would not have been a big problem, if the enormous costs of maintenance was not an issue but my professional experience shows that this particular circumstance is the main road blocker of digitization. This is the reason why very often companies invest into the infrastructure domain and

while not having much influence on driving success. Mainly because the effectivity cannot be ensured only having technology without a strategy of business transformation.

Actually it is quiet reverse the case, the effectivity and usability of even fastest- & most modern piece of technology will be not persuasive if there is no "business ecosystem" there and e.g. the management of data storing & processing is not done correctly. This among many a key reason why such ecosystem base digital strategy of e.g. CEE legacy banks massively fail to find a path to success.

2.3. Solution approach: Implementation of data management & governance in a bank

Data Governance [16] refers to the overall management of the availability, usability, integrity, and security of the data used in an organization. In times of emerging technology solutions creating data outputs nearly from every pipeline, it is one of THE important and fundamental issue in evolving business transformation.

One important goal of Data Governance is to maximize the company's value of its data assets. Within this context, data may not only generate a business value if it is being used but has also am immense influence to internal topics such as processes optimization and above all automation/digitization. One of the positive aspect/added value of Data Governance is e.g. the democratization of data access for all relevant business departments.

With the advent of data lakes, Data Governance has become more relevant than ever since such technologies more and more are used to consolidate several existing legacy or redundant systems within ecosystems of traditional Banks in Central Europe.

2.3.1 Implementation scope & problem statement

Since end of 2016 one of the biggest Austrian banks has been trying to add a Big Data Analytics capability to its current data ecosystem because it had many legacy systems in operation with nearly no interconnection beyond day-to-day usage. Historically seen most of business areas are creating up-to-date data within their "data silo" and there has been no room for using those data for e.g. predictive selling etc. models.

To enable a variety of predictive, descriptive and prescriptive analytic insights on the data assets of the organization, the targets of migration has been broad ranging, including improving customer engagement, enhancing next best offer

based programs, reducing customer churn and enhancing financial risk management.

To this end, the bank had engaged resources to create it´s future "Big Data based Digitization Roadmap" to define the architecture blueprint and next steps for this advanced analytics capability. The major focus have been lying on following points:

- Consider could-based managed services for the Big Data Lake environment. This is to reduce total cost of ownership (TCO) and deliver on the expected return on investment (ROI) the bank grows in-house capability to enable them to be self-sufficient.
- Establish a Data Governance team / structure to drive the Big Data program of work, involving key participants from the business, IT and the data office.
- Implement the recommended Data Governance processes and associated security protocols and access restrictions to ensure data is protected and handled only by those who need to have access to it.
- Develop skills in open source tools, both within the analytics functions of the business and within IT.
- Undertake Agile training and adopt agile methodologies for delivering the Big Data program to align with the business ecosystem
- Implement ROI driven Use cases within a time frame of 3 years

A series of deep dives, workshops and meetings in which the team conducted an assessment to evaluate current state data handling and practices against the above capability assessment model of Teradata. The model defined best practices for each of the capability areas in a Big Data Analytics & Digitization programs. The capability areas are usually structured around six top-level components - Strategy, Data Management and Governance, Services, Support, Operations and the overall program. One of the major areas detected by this model was e.g. data management & governance within the company since there was nearly no- or bad quality business knowledge, documentation of meta data [19] across the organization. Areas as marketing or commerce had built data silos of their own and had massive latency in providing own data to other departments for data-driven innovations (e.g. automated cross sell, predictive modelling and fraud detection).

As it is a key capability for any data-driven digitization competency, immense resources were invested in this topic to solve the issues. The Data Management and Governance capability included the following functions:
  - Technology governance
  - Data quality and metrics
  - Access controls

> Data Governance
> Metadata management [19]

2.3.2 Key findings of assessment

Technology Governance

Big Data technologies and automation/digitization capabilities had to be properly matched to business value and requirements. Data and technology governance was the right way to ensure that the right technologies are implemented for the desired outcomes. (Business & data ecosystem approach as mentioned in previous chapter) Standards and policies had to be established to provide guidance on when a new technology should be used and when it shouldn't be used.

Data Management & Governance – Technology Governance	
Best Practice Capabilities	> Adopt the right technology for the desired outcome. > Support incremental adoption of Big Data technology and concepts. > Understand and abide by requirements of all technology licenses, including open-source. > Big Data Technology Governance is implemented at the enterprise level (not departmentally or geographically).
Banks state in 2017	> The IT has chosen Hortonworks for Hadoop platform provider for the enterprise > Knowledge on Hadoop and Big Data in IT department is high however it is limited in coverage breadth and depth outside the IT. > The bank is seeking to adopt the good practices around the Data Governance
Recommendations	> Incentives and benefits for using systems in the correct way should be communicated broadly throughout the organization (both IT and Business). > Keep the current Technology Governance processes in place

Data Quality & Metrics

Data Quality starts at the data source. Although the Big Data platforms usually won't only create new data, they consume a significant amount of upstream data. In addition there would have been transformation rules embedded in the data transformation applications. So there should have been a proper processes, roles and tools enabled in order to manage and monitor Data Quality.

Data Management & Governance – Data Quality & Metrics	
Best Practice Capabilities	⟩ Define and document schemas for all ingested and derived data, optimizing storage requirements and access efficiency across all known use cases. ⟩ Use Data Catalog, when possible, to store and maintain schemas. ⟩ For long term archive of structured data use a self-describing file format (e.g. Avro, Thrift etc.) ⟩ Monitor data on a regular basis to ensure upstream systems are adhering to requirements. ⟩ Establish Data Quality metrics and share with all stakeholders. ⟩ As validation and normalization routines mature and stabilize, move them upstream into the data ingestion / transformation process. ⟩ Generate reports to create an auditable trail for accuracy and overall improvement of data quality.
Banks state in 2017	⟩ Most of the data that lands on Operational Data Lake comes from High Trust systems, that's why there are no defined quality checks and data profiling processes. Most of data quality checks are defined in Data warehouses and core banks system ⟩ Central Data Catalogue is currently not defined ⟩ Data that is ingested to Operational Cluster is treated with 'harmonization' processing which ensure that stored data has most unified structure as possible. ⟩ The bank is not currently measuring and reporting on any data quality metrics.
Recommendations	⟩ Establish Data Quality metrics and regularly communicate to all stakeholders. ⟩ Engage business users on the types of DQ issues and establish a DQ team to address the issues. ⟩ Establish a continuous monitoring / review process for Data Management & Governance to ensure upstream & downstream systems are compliant and adhering to requirements. ⟩ Implement the Data Quality Checks and data profiling processes while data will be copied Operational to Analytics cluster ⟩ Establish Central Data Catalogue and Data Dictionary which will be available to all stakeholders of the Big Data Analytical Environment

Access Controls & Security

Access Controls & Security in a Big Data & automation platforms can be very challenging due to the lack of standardization around security components offered by the major distributions even today. The appropriate level of security must be implemented based on the content of the data, it's sensitivity and corporate policies.

Data Management & Governance – Access Controls & Security	
Best Practice Capabilities	› Determine required granularity of access control. › Use cluster perimeter security and access API security whenever possible to achieve security goals. › If fine-grained security within the cluster is necessary, carefully evaluate offerings within major distributions: › Define User groups (e.g. analysts, data scientists and partners). › Define Data types (e.g. raw logs, testing, aggregates, customer, and product). › Specify User-specific access limitations for each data type resulting in a matrix of data types vs. user types with access eligibility. › Implement appropriate processes for each user type, data type and environment. › Prepare user information and training.
Banks state in 2017	› Operational Cluster has a limited accessibility. Only the IT team and Vendor maintenance team has the access to the cluster, › Operational Cluster is not kerberized, Access is limited on the network level. › The IT has carried out detailed Data Governance discussions with the Business regarding the Analytical Data Lake
Recommendations	› Setup the analytical environment with best industry standards of System and Data Security › Setup the auditing with Ranger, and automatic audit log monitoring › Define and achieve consensus on critical data governance for the analytical environment (GDPR)

Data Governance

Data Governance usually involves developing roles, processes, standards and policies on how data can be used. It identifies ownership and usage of data, and specifies retention, archive and other data lifecycle policies. In case of the bank several findings were made and recommendations given :

Data Management & Governance – Data Governance	

Best Practice Capabilities	❯ Build a Data Governance practice within the CoE with responsibility for:
	    o Evaluating whether potential new data to be ingested into the Big Data platform has particular security or privacy restrictions.
	    o Assigning appropriate data ownership and stewardship.
	    o Defining, documenting and communicating data usage policies.
	 ❯ Integrate the Big Data Governance approach within the business team.
	 ❯ Include within the Big Data Governance team active involvement from both Technology and Business.
	 ❯ New data types (e.g. photos, web logs, and sensor data) and new metadata structures are standardized across groups and systems.
	 ❯ Exceptions to data standards are resolved without impact to projects or the Business.
	 ❯ Mindset: Data Governance is an enabler to projects, not a hindrance.

Banks state in 2017	❯ The bank wants to setup a Multi-tenant Environment ❯ At current state the bank has set up the Big Data Governance Team nor formal Big Data CoE ❯ Most of the data that will be needed to implement use cases are already ingested to the data lake or planned to land there in near future ❯ Big Data Governance policies has not yet being developed
Recommendations	❯ Create a wiki to describe the data in the governed zone: how data is partitioned, how it is serialized, meaning of each field, source, ingestion frequency, processing pipelines. ❯ Formally establish a Big Data Governance team with clearly defined SLAs, services and roles ❯ The consequences of poor Data Governance should be well understood across all stakeholder groups. ❯ Define a data steward role to enforce Data Governance and quality of the data that are ingested and processed by the analytical environment

Metadata Management

Metadata Management is the end-to-end process and governance framework for creating, controlling, enhancing, attributing, defining and managing a metadata schema, model or other structured aggregation system, either independently or within a repository and the associated supporting processes.

Data Management & Governance – Metadata Management	
Best Practice Capabilities	❯ Use data catalog to manage data schemas and partitions whenever possible. ❯ Index ingested data with versioned schemas in situations where structure of ingested data varies by source or time. ❯ Provide mechanisms to abstract away schema variations to supply a single, logical, model for the underlying ingested data. ❯ Use of self-describing data structures (e.g. Json, avro) where appropriate. ❯ Key data element definition and business logic should be well understood and maintained across the organization.
Banks state in 2017	❯ The bank was for data catalogue with departmental ownership (describe & steer); but access on data should be provided cross-departmental. ❯ In current operational cluster most of the data that is stored on Data Lake is registered in catalog
Recommendations	❯ Operational metadata needs to be captured at all stages of data transformation ❯ Define the end-to-end metadata and lineage tracking goals ❯ Investigate various techniques for flexible schema management and identify which ones fit best with the nature of the data, tools and development processes. ❯ Integrate metadata management tightly with the Data Governance procedures

2.3.3 Organisational set up

Data CoE

In most organizations there are a couple of obstacles to overcome to embrace the new ways of doing business data-driven approaches open up. Even today there are a lot of companies lacking understanding how to use Big Data and Analytics to improve and digitize their business. If it is done it is not performed throughout a central organisation the management bandwidth is often

insufficient due to competing priorities within the departments. The departments internally often lack of skills in line with the business. Without the backing of a Big Data CoE even promising initiatives might suffer from inadequate executive sponsorship. The culture of many organisations does not encourage the sharing of data across departments. This leads to an inability of getting data, due unclear ownership and concerns with the usage of this data.

To support a modern data ecosystem, which das a proper Data Governance structure layer around it, several disciplines needs to come together. The data CoE should become the hub for data-driven initiatives across the organisation providing the skills, methodology and governance to integrate the enabled 'spokes' of the organisation and provide tangible value to organizations. Regarding general it tasks about network operations, etc. the data CoE does not replace a regular IT department.
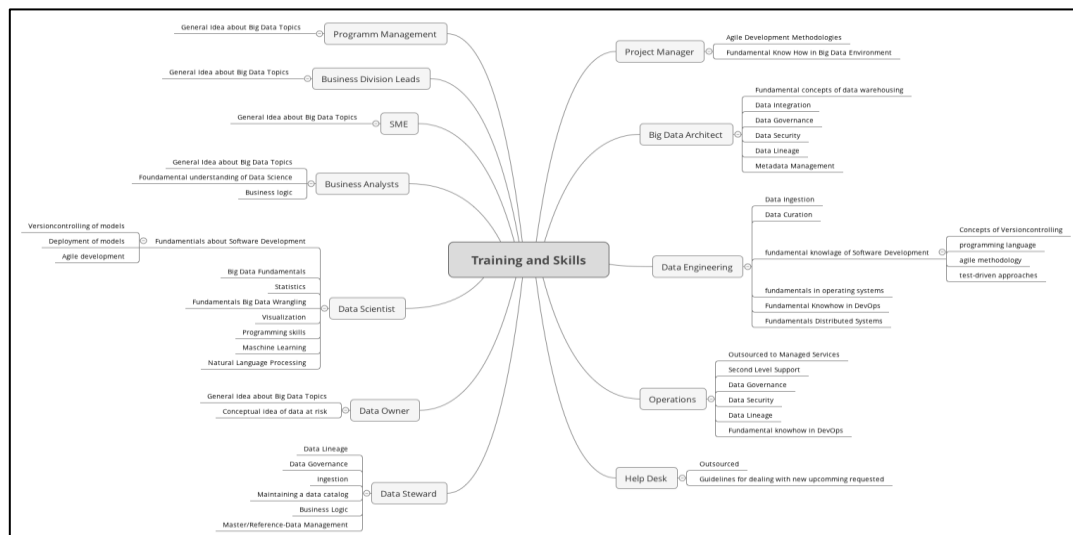
Picture 20 © Alin Kalam

All these tasks are the foundation for a working IT environment and needs to be performed by regular IT, changes

impact the Big Data environment needs to be done with notice of the data CoE or in collaboration with the whole company, since it is the single point of contact to all data-driven business belongings. Such an approach was also taken by the related bank and the major areas of the data CoE were:

> Strategy
> Program Governance
> Data Management
> Operations & Devops

Roles

In industry this section deals with roles necessary for successful implementation of use cases at enterprise-scale. Roles and skills are split into technical and business related roles. All descriptions are organization agnostic and in following picture below it is illustrated how Data Governance roles in term of "Single Point of Contacts" were implemented within the new data-driven organization

- Definition of **roles and responsibilities**
- Enforcement of Data Governance **policies**
- Definition of **processes** within the company
- The collection of specific data objects complies with the **legal and regulatory requirements** (PII, security & etc.)
- Creation of **Data Dictionary with value, priority, owner, metadata etc.**
- Building of a **security & access matrix for data consumers**
- **Monitoring** of regulatory & compliance conformity
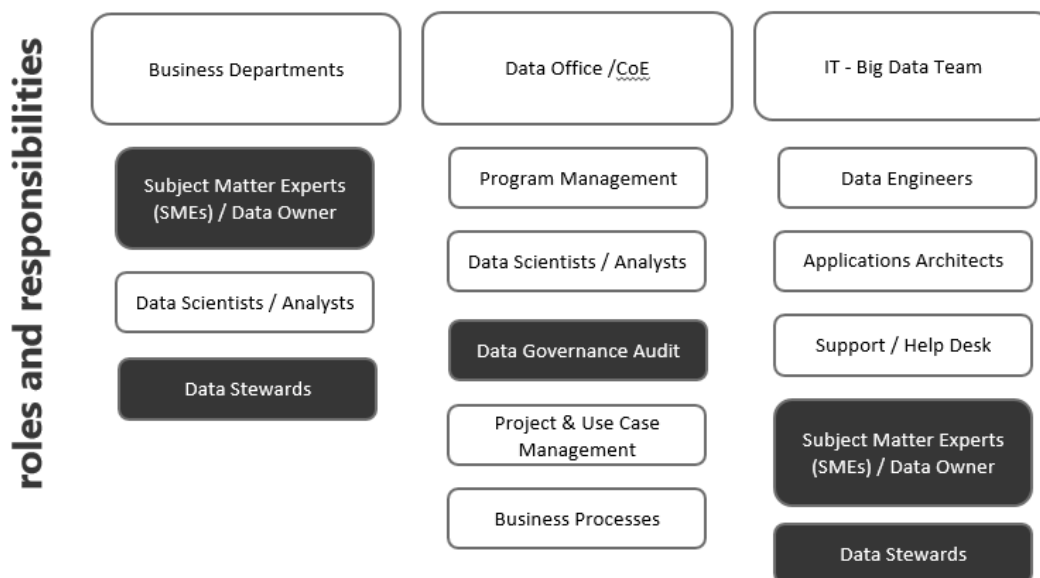
Picture 21 © Alin Kalam

2.3.4 Technical approach

As the actual architecture setup is a matter of market leadership and innovation of the bank, it was not possible to get detailed insight into it. Nonetheless in the course of this work it was possible to cooperate with usual market providers of such solutions like Hortonworks, which was beside own experience and scientific research paper of ScienceSoft [17] a major key factor to illustrate a potential solution. It correlates directly with the Data Governance concept of the example of the bank described above.
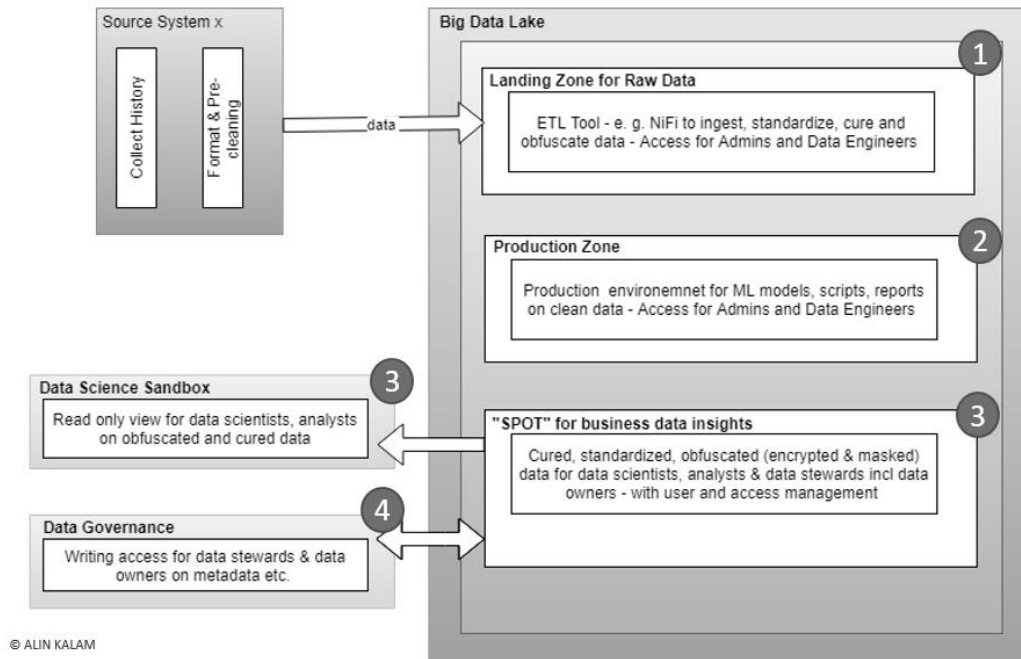
Picture 22 © Alin Kalam

The reference Data Governance architecture illustrated in picture above subdivides the ecosystem in four major zones:

1: Landing Zone: it is the landing container of raw data, that gets ingested from legacy systems or other solutions e.g. open data or external data. Raw data is also known as source data, primary data or atomic data.Usually this area is not not accessible for business out of security and usage matters.

Raw data refers to any data object that hasn't undergone thorough processing, either manually or through automated software solutions.

2: Production Zone: it is the container where curated and standardized data from raw data container is copied. This area of such a data ecosystem is also not accessible for business users, since it is only used to deploy and productionize e.g. data models. Some examples would be:

- Predictive, perspective etc. selling data models
- Fraud detection
- Real-time customer satisfaction etc.

Only in exceptional case data scientists/analysts have reading access to this zone and usually they are also "SPOC" or "Data Owner" at the same time.

3: SPOT/Data Lab: This is usually a virtual R&D zone where broader range of business user can read and right on the virtual data without any impact on production zone.

The Data Science Lab is an environment where a data scientist or other data consumers from business are free to develop complex models on production-quality data and other data. In times of cloud computing this lab zone is often implemented as a separate cluster, may be spun-up and spun-down as required in order to provide appropriate resources without affecting other production processes in production zone. Cloud environments such as Microsoft Azure and Amazon AWS enable the on-demand nature of Data Lab resources.

Data Labs are typically partially populated with data by data scientists on-demand as well as being populated by Production or copies of Production data. The Lab typically is the source of data models that can be deployed to a Data Lake. Depending on analytical use cases, Data Science models will take form of either scripts (often in R, Python or Scala). These models are then integrated into data pipelines running on the production zone.

4: Data Governance Zone:
In process of building a data-driven automated/digitized business, it is indispensable to explore data from other company units for feature engineering, statistical model creation, testing and, ultimately, for model-pipeline creation and testing purposes. As mentioned, legacy banks usually have mainly their core data set (master data) well documented. Beyond the master data tier, researches and scientific papers as "Metadata Use in the Commercial Banking Industry" by Joan Starr [18] "& Modelling metadata in central banks"[19] by David Bholat prove that there is a huge lack in terms of collecting, interpreting & using "data about the data" (metadata).

Business and operational metadata is often one of the more under-prioritized areas of data implementations. In times of boundless collection of quantitative data through out every available touch points with customer, it is highly recommended prioritizing the collection and management of metadata for data-driven transformatio. Types of metadata may include:

- Input data (data description, size, timestamp, checksum, type of data [txt, csv, zip, rdbms, etc.], ownership, schema, feed preconditions, SLAs, etc)
- Operational metadata (explains how datasets are created and transformed)
- Data confidence measure (tracking of freshness, completeness, etc.)
- Data quality
  - Accuracy
  - Completeness

- o Consistency
- o Freshness
- o Uniqueness
- Provenance (tracking of relationships between dependent feeds, change datasets and operations)
- Application specific metadata (information captured about certain applications and tools that tie into metadata store)
- Security-focused (data owners, security groups)

Beside a well-documented core master data set, metadata builds more and more enormous added value to business automation/digitization of e.g. processes or even fine-tuning of data models. According to NISO (National Information Standards Organization) Metadata is key to ensuring that resources will survive and continue to be accessible into the future [19].

The Data Governance zone is a reading & writing access zone for Data Governance related roles e.g. "data owner", "data steward" etc. on metadata. Ever since the interpretation of metadata is mainly a manual qualitative process, mostly depending on human, which creates data quality issue depending on the accuracy of inputs. Nonetheless since years there has been outstanding progressions in this field where more and more machine learning algorithms help detecting failure, quality issues and interpreting data sets fully automated. Provider as Colibra, Informatica, Alation & many more currently offer metadata management tools for Data Governance purposes. One huge added value of those solutions are, that e.g. also master data sets can be "curated" and treated the same way to create an universal enterprise "Data Governance Hub" for a huge quantity of business users.

2.3.5 Implementation of Data Governance in day-to-day business

By implementing such a data management & governance software solution incl. data catalogue, interpretation etc. suddenly there is an enablement of cross-domain data usage without generating manual effort, which again helps saving a lot of time and money.

In case of named Austrian bank above, a quite innovative and niche solution of the provider Alation [21] was implemented to the data ecosystem. According to it´s own website Alation is a data catalog & governance tool where everyone within an organization can find the data they need to collaborate. It automatically indexes data by source. Below a reference illustration:
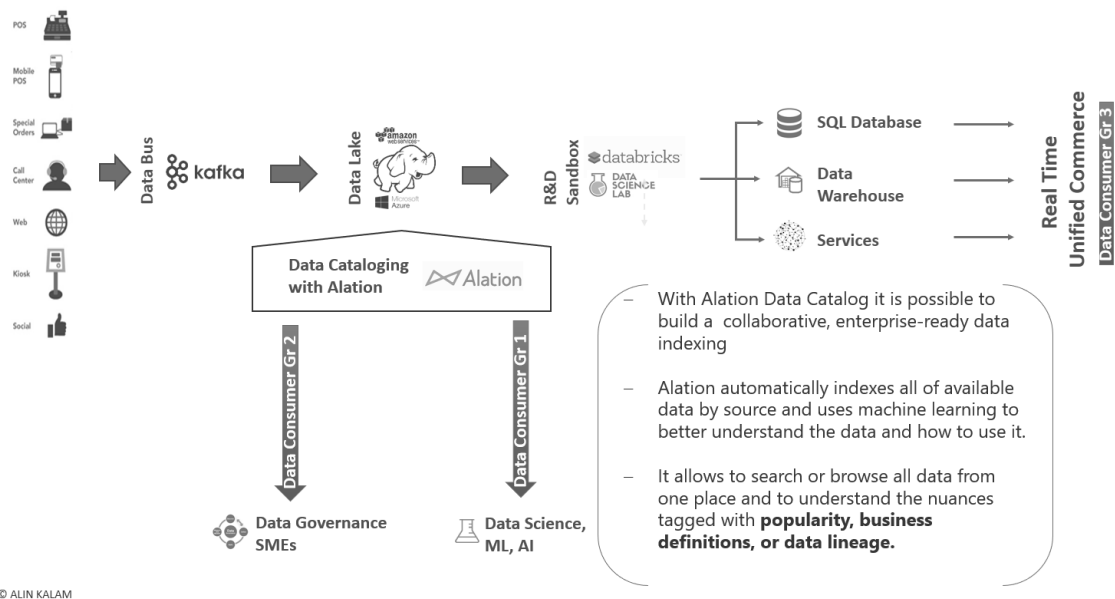
Picture 23 © Alin Kalam

It also automatically gathers knowledge about e.g. the frequency of usage. Like Google, Alation uses machine learning to continually improve human understanding.

To interpret the business context of the data, most organizations have evolved a complementary and ad hoc collection of wikis, email threads, and chat sessions. These unconnected tools are used to identify subject matter experts (SMEs) and to communicate about data. Sometimes more formal documentation is produced by a data steward or data curator in a technical repository, or by a developer in the source code. Even in the most advanced analytic driven direct banks of Central Europe often rely on physical notebooks to store business meta data [19] notes.
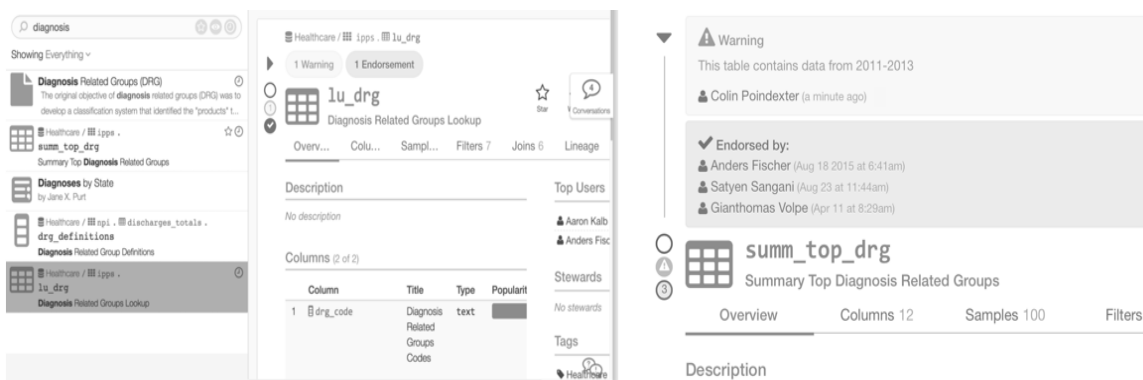
Picture 24 © Alation

2.3.6 The mid- & long-term impacts of Data Governance on business

This Data Governance approach is a is digitization case on its own! E.g. because Alation observes and analyzes the relationships between people and data. At a time where the size of stored data is increasing at a monumental rate, the speed by which machine-based algorithms can capture these relationships is crucial. Hiring more people to manually document data is no longer a scalable solution. The more organizations automate the collection, cleansing, preparation, documentation and dissemination of data, the faster organizations can derive value from the data.

Such an approach can be used from business & IT departments in:

- Avoiding data duplication,
- Raising data quality through automated detection & harmonization of wrong data,
- Connect data consume in business with related roles, ensuring ownership and etc. rights
- Enabling fast cooperation between previous data silo based departments to accelerate data-driven innovations

2.3.7 Learnings, improvement areas & summary

According to several interviews in business departments of the named bank had a few major challenges in cultivating Alation as central Data Governance tool, mainly because it does not cover up entire spectrum of

Picture 25 © Alin Kalam
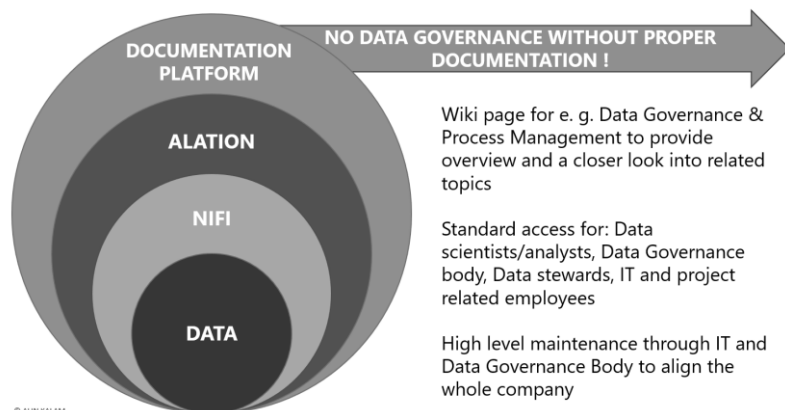
services to solve legacy issues such as latency between data providing systems or even formatting issues of several data types etc.

In Addition to that according to many the internal knowledge transfer should have been the process responsible for sharing perspectives, ideas, experience and information and for ensuring that these are available in the right place and the right time while providing trainings for business users.

This function usually enables informed decisions and improves efficiency by reducing the need to rediscover knowledge. It also includes a central

repository for all knowledge captured in the CoE. That's why best-practise analytical models, technology decision factors and reusable workflows / practises, among others, should have been stored in the repository for organisational reference of the transformation phase from manual to automated Data Governance approach.

At most, there has been the opinion that an online documentation platform such as JIRA, CONFLUENCE, Wiki could complete the technical tool stack to cover up missing functionalities of Alation, which is e.g. putting historical relevant illustrations, data samples (not in production) on the frontend to clarify the usage of several data sets. Without governance a wiki might turn from a knowledge hub into a knowledge dump. A "two set of eyes" approach will ensure all content is going through an informal review.

CHAPTER 3.0

3.1 RESEARCH AREA 2: Impacts of data-driven & ecosystem-based digitization on management, strategy & organization

Hypothesis 2: A data-driven management e.g. enabled through Business Intelligence and Data Science adds additional strength to business and fastens the digitization process of big companies:

The impacts of data-driven digitization on management and company strategy are immense when it is done right. In center of all digitization approaches must be the company mission, vision & goal and that should be the driver of every innovation & optimization processes.

One of the best positive aspects of data-driven mindset is the unchained "democratization" of data and the right to use it regardless of whether it belongs to your department/ business area or not. The main aim is to generate knowledge based added value for the company. Tus, the over-all challenge to build s fine tunned technical ecosystem coupled with concepts of Data Governance and modern data processing methods only does not solve all problems of existing companies. As mentioned in chapters above, it could be seen as major enabler in driving innovation with effects of e.g. cost reduction by optimization, automation or even improved customer interaction. Nonetheless many interviews and research on existing projects of Lufthansa, Austrian Airlines, Raiffeisen International shows companies often miss to carry data-driven thinking into management or decision making tier.

As mentioned above there is one major factors that must be taken care of before even thinking of saving- or earning more money through further

optimization and automation, that is all about "transforming management board & decision makers" to data-driven in the 1st place. Tus, it must be mentioned, that management boards of big companies usually have been working data-driven for decades, what has changed is the market circumstances. In many fields decision makers are forced to react and decide faster than ever, because markets but also customer bases change very fast. At the same time those are the key revenue generator that are very important to the business. The usual approach of detailed focus in western economy has been the financial KPIs such as "return on investment", "market share", "return on equity" and etc. Nonetheless current business circumstances with rapidly increasing digital market share (online, omni-channel etc.) leave no room for only finical focus.

Mainly because often data-driven digitalization is all about building or extending the technical ecosystem supported by profitable business case related motivations and foresights into the tech driven company we want to become rather than only investing money in technologies, hypes or buzzwords with ROI just to be a part of something. As a profitable company in early 2020s optimally we should have an open source based flexible extensible IT fundament consisted out of many components and sub-systems & please make sure that you drive innovation in that particular destination. The data-driven management & strategic foresight can be seen as the spearhead of these efforts. In other words, it is very likely that many jobs will disappear from the market through further digitization or automation of our industries, which often leads to the assumption that digitization/automation means only cost saving ! This specific assumption is very risky in many ways, because e.g. it might be possible to optimize offline retail by merging/relocating branches but on the other hand for sure companies might have to invest in digital strategy, decommissioning legacy IT systems, redesigning business processes & customer journey or even build therefore extraordinary sophisticated modern branches where the customers will find indispensable physical touchpoints and services offered. This is very often easily said than done in a times where customer churn happens much quicker than any time before in the history and therefore this is a true challenge we all have to find solutions for. If right steps are undertaken and investments are made thoughtfully by having ROI generating "low hanging fruits", it will be possible to push innovation and company culture forward.

Companies e.g. Amazon pretty much shows the world of doing things right and investing into an ecosystem of technical key components and services around them to enable "others" to drive data-driven innovation. Therefore a big chunk of Amazon´s revenue is generated by their loyalty program "Amazon Prime", which contains the whole idea of ecosystem & data-driven thinking. Such strategic mindset and initiation of right projects and right investments in right focused area of a data & fact driven future is the open secret to success that starts and ends with such enabled management tier. In

current chapter we will get more into the aspects of such data-driven management, because it might be one of the major pioneering success criteria of digital transformation and the main reason why so many companies fail to succeed.

Data-driven approach might be the core of digitization but data-enabled management is therefore can be seen as the nutshell around it!

3.1.1 How data-driven is connected to management

Digitization or digital transformation is ongoing process of introducing computer-based solutions that convert analog data into computer-readable or digital format and at same time improve the efficiency of organizational process. Examples of such transformations are:

- Exchange of paper based boarding passes with electronic boarding passes
- Exchange of hand written and paper based forms with computer based electronic forms
- Exchange of paper based monetary exchange with electronic payments
- Exchange of various MS Excel spreadsheets with one centralized collaboration software
- Exchange of oral based information exchange with computer based live dashboard
- etc.

Reason is simple. Analog data can be scattered throughout the Enterprise and typically suffers some loss of quality each time copied or transmitted. On the other hand, digital data, in theory, spread indefinitely with absolutely no degradation and is usually stored centrally in one system. Furthermore, data mined in disparate silos is time consuming, it undermines the quality of decisions across the enterprise, increases risk, reduces the security of corporate data, reduces efficiency, and drives IT costs up. All that leads to the conclusion that digitization is of crucial importance for proper data transmission, storage, processing and analysis.

Management

Business Intelligence	Data Governance
Comprises the strategies and technologies used by enterprises for the data analysis of business information.	A set of processes for defining, implementing, and enforcing data policies to meet a company's mission or to achieve specific business objectives.

Operational Level
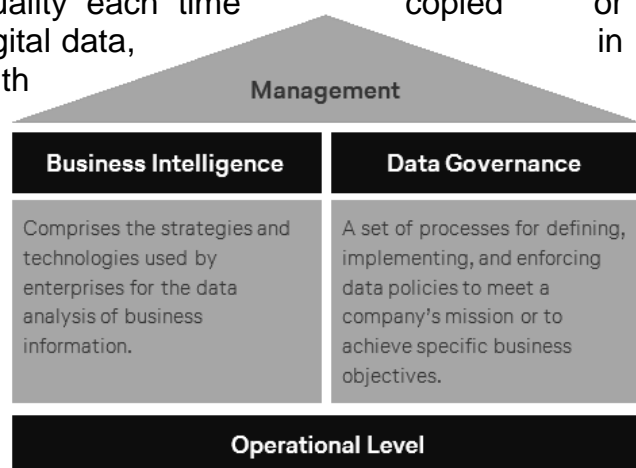
Digitization should serve decision preparing process. It is digital transformation of manual decision support system and enables executives and managers to focus more on decision-making process rather than on decision-preparation process. As such, it is a subset of overall digital transformation and major legacy companies often make major mistake by labeling "data-driven" as operational component only.

3.1.2 Why data-driven management tier is a "must have"?

Both in the business departments and in top management, many complex questions with synergy effects between key components must be answered to make important decisions. This ranges from issues such as:

> "What added value do our strategic initiatives create?"
> "Which similar-sized companies have e.g. anticyclical seasonality?"
> "Where is the net value of digital strategy & investment in data ecosystem?"

With the right mindset, the right data and the right tools, companies can derive measures as early as possible, provide concrete answers in a short time and measure the actual value creation after the final implementation of the measures, which are unlikely previous times beside only financial KPIs a lots of other key facts and measures.

An example of the simple evaluation of company data is the profitability calculation of omni-channel. It takes into account the data of each channel sold to obtain an overview of the profitability of individual products or customer status levels within the business field. Strategic data-driven capability in management is therefore a key success factor in controlling any business and maintaining competitiveness in a rapidly changing market environment.

Responses to strategic issues and causal relationships can be provided in real time and in a simple, visually attractive form, thus strengthening the agility of the company. The long-term goal of such initiative is to create an environment in which one ensure an innovation-promoting culture of analysis and a continuous learning process. Thus, every employee should be encouraged to look outside the box and to go on a "search for clues" in his area. For every company, this is another step in the direction of digital transformation.

Nevertheless this is not the only success criteria of becoming successful. The modern approach of "democratizing" data and the usage data-driven insights throughout whole organization doesn´t often work well because of traditional top-to bottom hierarchies, which is often cemented by legacy processes. In other words, data-driven digital transformation & management demands adjusted organizational structures and also process improvements, because

digital transformation of business does not only take place at one particular area but much more sets foot prints in nearly every business areas.

Digital transformation
As a permanent, structural change based on technological trends

Cycle of digital and structural change

Market consolidation

6

New technological trends

))NFC))

1

Growth of new players

5

Data-driven Management

2

Google

New competitors

⊙ Lendico   ⊛ NUMBER26

4

3

Proven revenues sources under pressure

Change of the customer behavior

Source: KMPG, Digital Transformation
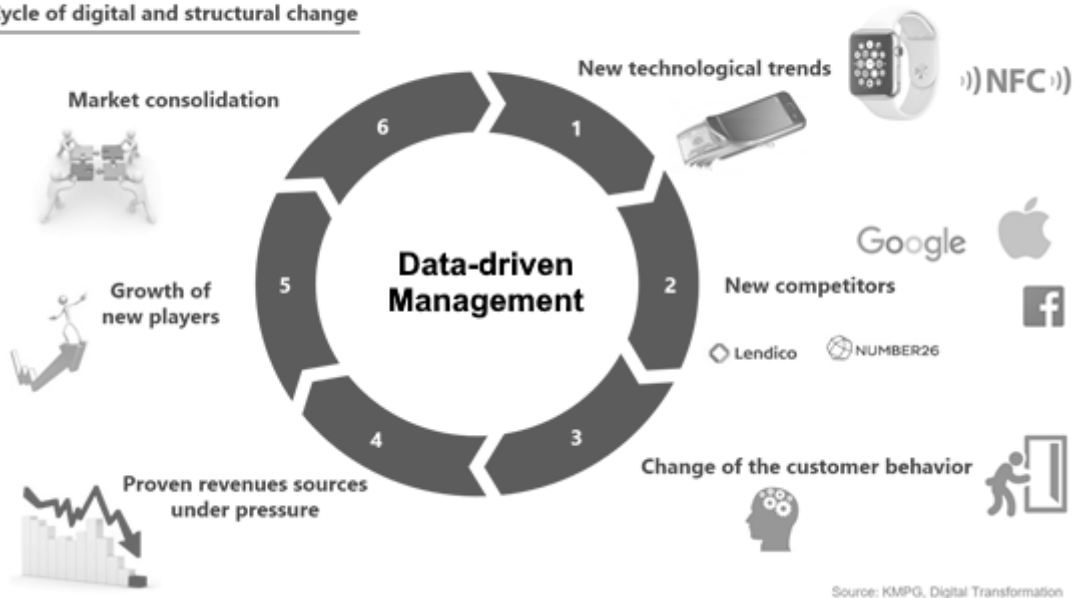
Picture 27 © KPMG, BAWAG

3.1.3 Detailed market analysis – Spotify, the figurehead of agile approach and a story of data driven flat hierarchy

Beyond ecosystem based transformation of companies to data-driven management, there are several important areas that are major influencing domains such as development of:

> Processes
> Organizational structure
> Company identity & customer orientation
> Working culture

At this part of the thesis, a major focus will be projected on current market situation and key success factors that are complementary and very important to every data-driven digitization approach. Below the most important ones are mentioned

3.1.3.1 Process transformation – the unspoken kingmaker

In particular there is either usually a lot of room to optimize- or even rethink and redesigning of existing processes. By going through data-driven transformation process, companies often find themselves investing a lot of time and effort into redesigning of processes.

An example could be the implementation & cultivation of such technologies as "Data Lakes", where lot of unfiltered & massive amount of unstructured and structured data can be ingested. On one hand this approach might enable companywide data consuming and innovation to build e.g. new products, but on the other hand even though Data Governance & management is in place there can be massive data leaks if the process of consuming, writing or copying data out of the system is left behind. Without well thought-out and integrated automated processes, this kind of innovations can become very dangerous, because General Data Protection Regulation (GDPR) requires
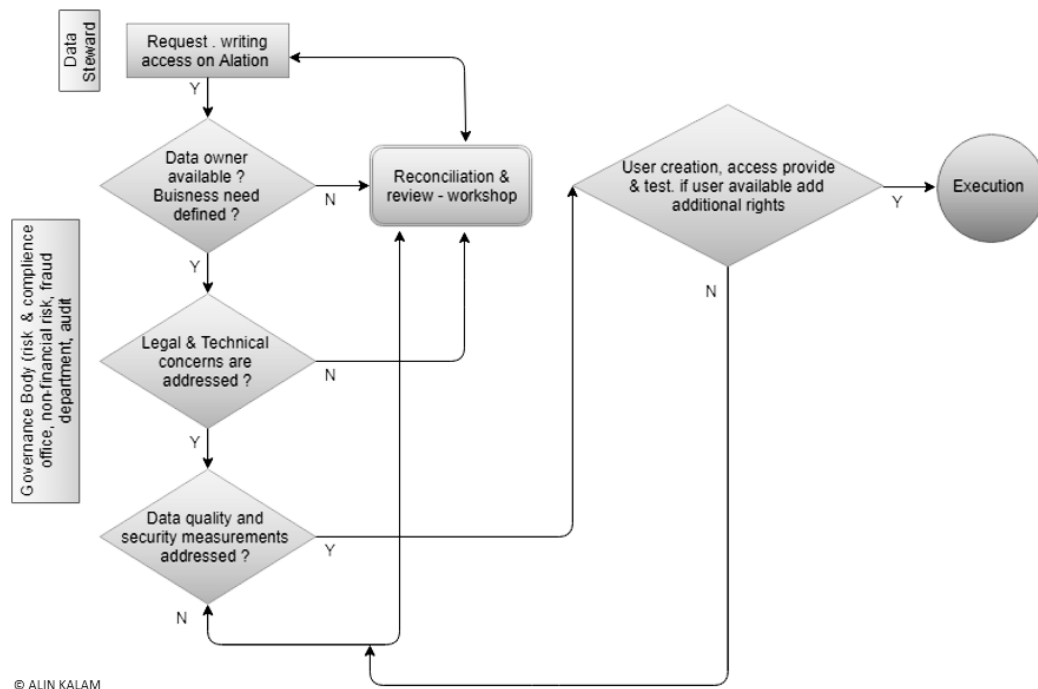
Picture 28 © Alin Kalam

companies to ensure lawful data collection, data processing, etc. The earmarking of the data processing is one of the core elements. Personal data must henceforth be collected for specified, distinctive and legitimate purposes and may not be further processed in a way incompatible with those purposes. In particular, the retention period for personal data must be limited to the absolute minimum required. Understandably, companies also need to protect adequate security of these data through appropriate technical and organizational measures.

In 2015 Gartner predicted that by 2017 more than 60% Big Data migration projects will fail to go beyond piloting and experimentation and will be abandoned due to lack of business case and proper use cases. Even though the over-all maturity of technology is given companies tend to create "data swamps" tend to be formed by copying data from all areas into the new data systems without cleanup, standardization, anonymization / masking or even authorization domain. [22] The risk that personal data is visible for non-related users in plaintext or misuse is of course inevitable and according to the GDPR may cost 20,000,000, or (if higher) 4% of the annual turnover!

Coupled with ethical standings of digitization and privacy/protection needs which are not only regulatory but also sometimes moral barriers, it is important to follow a holistic approach. In next chapters such areas as "people analytics" will demonstrate the thin line between business needs and ethical/ moral stand points. This will mainly underline the fact that even today in our automated and optimized industry in western world, majority of processes within companies have huge amount of touch points with human, regardless if it is through interaction between man and machine or automatically generated.

Major challenge
Usually in such companies there is a huge amount of working processes existing that are optimized with up-to-date technical frameworks. Most of those processes create the modern day equivalent of "logbook entries", which detail exactly which activities were carried out when and by whom. If maybe a purchasing process is started in a software system, every step in the process is indicated in the corresponding data tables. Similarly, CRM systems, ticketing systems and even legacy systems, record historical data about the processes.

These digital traces are the byproduct of the increasing automation and IT support of business processes. According to the process mining book [23] the systematic analysis of digital log traces through Process Mining tools offers enormous potential for all organizations that are struggling with complex processes. Through an analysis of the sequence of events and their timestamps, the actual processes can be fully and objectively reconstructed and weaknesses can be uncovered. The information in the IT logs can be used to automatically generate process models. This approach is called Process Mining, which can be an immense added value to data-driven transformation of business. It not only solves the discrepancies between theoretical- & productionized real-life process that often differs
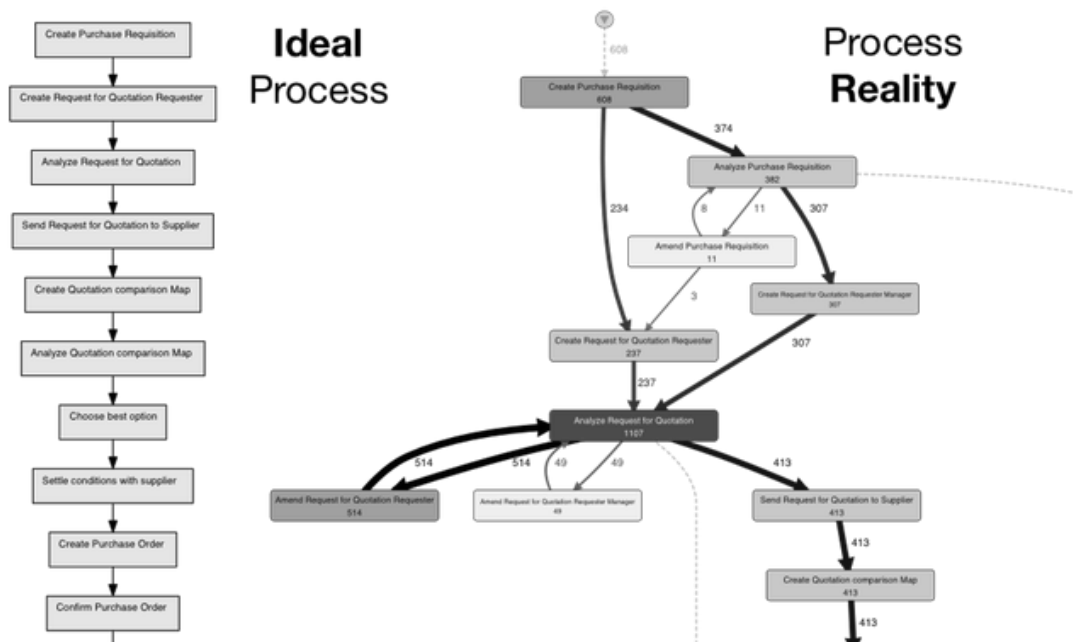
Picture 29 © Processminingbook

Especially when it comes to low-cost carriers within European aviation field, for years we see an unrestrained process optimization to e.g. save cost. A light house example could be Ryanair proving to many legacy airlines that this concept has a lot of potential and above all business case at the core.

A part of aviation, the producing industry in Europe & America can be seen as THE most optimized fields in our economy. But human history and experience level show that the feverish search for optimizable processes eventually comes to an end on the long run. Therefore it is not only important to optimize but at the same time companies have to transform their working culture and above all break old organizational structures to bring calm but also creativity inside the company.

3.1.3.2 Spotify – a market analysis on going new ways

Spotify has become a popular music streaming provider, well known for providing original and a limitless collection of music to listeners. It was launched in 2008 and has become a large company with nearly 1600 employees since then. Beside the implementing revolutionary Big Data ecosystem, modern analytics and statistics models to predict music taste of customers and many more things, they owe their success to their deeply rooted agile methodologies and the utilization of the Agile Scaling in their own way. This method is called **the Spotify Tribe**.

It is a never-seen-before success story of a company which pretty much has torn apart the legacy way of company, organizational structures to build an overall flat biracial organization.

The Spotify tribe
Spotify teams organize themselves among others in squads and chapters. Squads are the smallest units. Colleagues working across disciplines (less than eight in Spotify) work in largely self-sufficient teams in a clearly defined area of responsibility. For Spotify, for example, this would be "Search" or "Playlist Management". Colleagues are no longer tied to departments as they used to be, but become part of a chapter. The chapters have the task of organizing the exchange of experiences between employees with the same skills and tasks - for example, software developers. The goal is for them to learn from each other and, if necessary, to develop certain common methods or standards. But this can be handled differently in different projects.
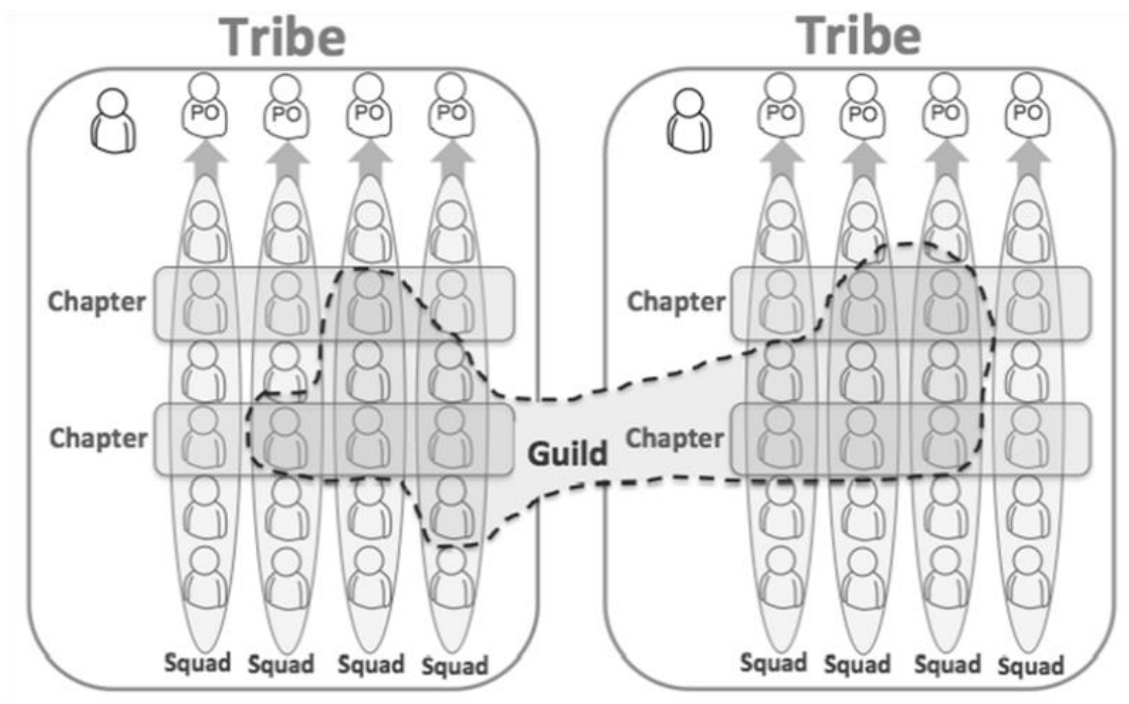
Picture 30 © medium.com

The combination of development and operation teams creates synergies and promotes smooth project handover. In surveys Spotify was able to prove that the employees are happier because of the flat hierarchy within and between organizational structures. This affects the entire work and the success of the projects, It is often the small things that make a difference because within the chosen constellation there is no longer talk about projects and subprojects avoids associations with traditional project management methods. It's good to use new terms. Unlike regular project organizations, the chapter functionality ensures that certain elements or practices and experiences are shared across the organization.

Implementation of agile in data-driven environments:

The ultimate foundation of the Spotify organizational structure is autonomy and trust within a flat company hierarchy. When there is trust, then there is ownership and accountability of the work done. Trust helps to create an environment where failure is seen as an opportunity to learn, innovate and change accordingly. This also uplifts individual morale and growth when working e.g. with customer data and can increase moral/ethic responsibilities. [24] The benefits could be::

> Few rules, easy to understand and quick to implement because often e.g. data scientists work with hypothesis and assumptions, which must be tested fast
> Short communication paths if necessary
> High flexibility / agility through adaptive planning, productionizing
> High effectiveness through self-organization and moral standards
> High transparency through regular meetings and backlogs
> Timely realization of new product properties or increments
> Continuous improvement process and optimization with equivalent team members
> Short-term problem identification
> Low administration and documentation effort
> Enhanced velocity of innovation

Despite the triumph of agility and related concepts real life cases show, that many companies have massively failed to implement such a working and organizational structure, because of a very common and wrong assumption that agility is all about cost saving and being faster than traditional project management or company structure. In research interview Mr. Stefan Bauer (Director of Talent Garden Austria) said his opinion on agility and data driven approach:
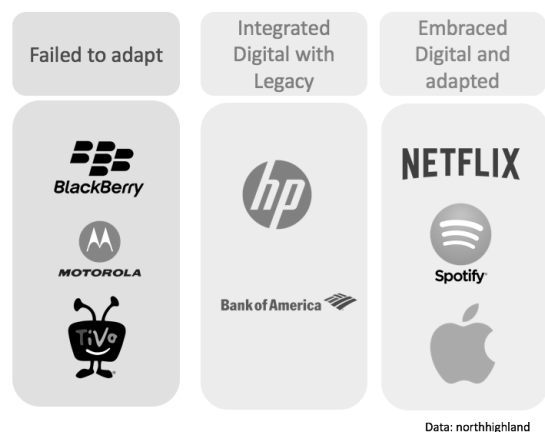
Picture31 © northingland

Data: northhighland

"At its heart agility means adapting flexibly to changes and reacting quickly to them - for example with innovations. An important element is as mentioned the self-organization. Teams that do not have to submit a thousand applications and work through hierarchies can act and react faster. It's also easier to innovate, nonetheless that should not automatically mean that things are getting done faster"

So it's not about introducing the same concept to everyone, but about understanding agility as a regulator that can be interpreted in different environments. Crucial here is the starting point: the management and corporate culture. Thus, the path from command-and-control to agile self-organization is unquestionably wider than the path from cooperative leadership to agile structures. The self-organization also has different levels. It begins with a semi-autonomous working group and ends in the complete decentralization & democratization which might become also sometimes contra productive. Some disadvantages could be:

> High level of communication and coordination effort
> Few concrete recommendations for action
> Lost time in "defensive" sprint planning
> "Tunnel vision danger" with focus on tasks only
> Difficult coordination of several development teams in major projects
> Potential uncertainty due to lack of responsibilities and hierarchies
> Potential incompatibility with existing corporate structures

Role of agile approach in terms of data-driven digitization:
Nevertheless, according to Mr. Stefan Bauer, his previous company BAWAG PSK group has been trying to adapt and implement agile concept from software development- to business departments across the company. According to him agile organizational structures worked within innovation driven areas as "Digital Banking" where it is the ultimate setup to drive data-driven and innovative thinking to build sophisticated customer experience and digital services.

By adapting their own tailored & unique agile scaling method has not only made Spotify achieve their goals effectively but also brought a shift in the mindsets of the employee in it. This part of the market analysis focuses on the

Spotify Tribe. Agile concepts as such are becoming more and more popular beyond IT and technology focused companies and will be beside ecosystem,
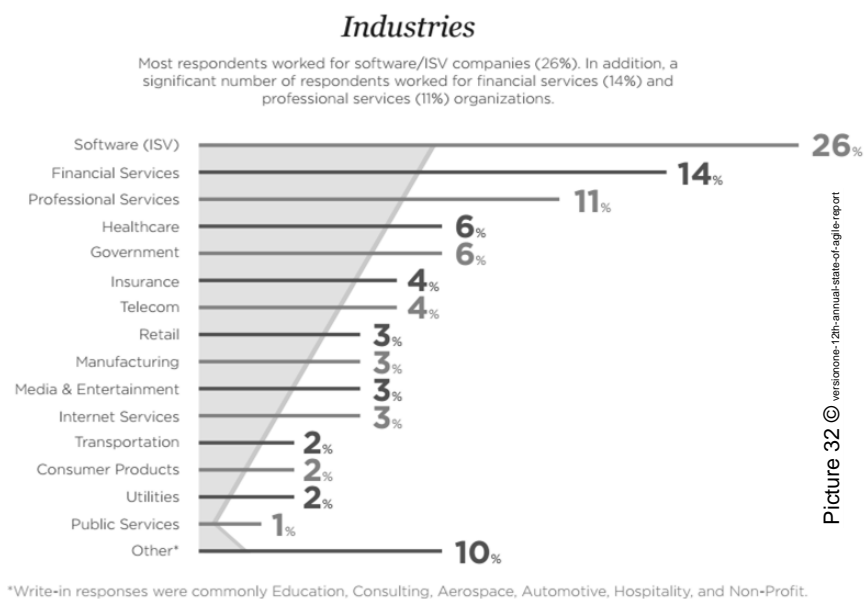
data-driven approach etc. one of the major success criteria for bigger companies.

According to the versionone-12th-annual-state-of-agile-report [25] agile methodic and organizational structures is has gained over 40% over the last 5 years.

Picture 33 © versionone-12th-annual-state-of-agile-report

3.1.3.3 Meritocracy in the workplace through data-driven digitization & agile approach

[21]In 1958, Michael Young coined the term "meritocracy" in his book, *The Rise of the Meritocracy*. He used the term satirically to depict a United Kingdom ruled by a system that favoured intelligence and merit above everything else. This would have been based on personal achievements of each individuals of the society. In times of emerging data-driven innovations, everyone in a company would have the right to express their opinions and are encouraged to share them openly and often. Those opinions are listened to and decisions are then made based on those that are deemed the best. According to Dr. Ansgar Allen, "once individuals are made responsible for their own meritocratic repositioning, a perfect distribution of abilities is no longer seen as feasible nor is it required" [22]. Within the technology industry the company Red Hat might be the ideal example of practiced version of meritocracy. A closer look on red Hat´s HR page confirms this assumption:

"Working at Red Hat is not a zero-sum game. You can't succeed and move up at Red Hat by yourself. Red Hat is a meritocracy where reputation is earned by how well you help others succeed. We succeed as an open source technology company when we create more open source winners. And we succeed as individuals when we help create more winners within our ranks."

[23]Although it is maybe essential to mention that, that a meritocracy is not a democracy. According to Management review article - "Achieving Meritocracy in the Workplace" of MITSloan, there is no "decision by consensus"; not everyone has a voting right. This is the key distinction of the meritocracy. While everyone does have a voice, some are listened to more than others because of the reasonability of the idea. In other words it would be a system where *best people and best ideas win.* Combined with agile driven approaches

meritocracy becomes more likely to step in organizations bringing new ways of thinking and leading, which is a very big part of data-driven digitization.

In chapter 5 it will be illustrated how e.g. people analytics drive meritocracy within organizations that are using new ways of data processing and interpreting to improve themselves and their employee performance.

CHAPTER 4.0

4.1 Scientific investigation: Data-driven digitization of European aviation sector

In order to substantiate the hypothesis of this research area, in following chapter one of world's biggest airline group Lufthansa will be investigated thoroughly.

To become the first choice for shareholders, customers, employees and partners in the aviation sector and to continue shaping the global aviation market as a key player in the future, many airlines have recognized consolidation, flexibility and digitalization as the key value drivers in the aviation value chain. Relying on own working experience, interview with decision makers and overall scientific research, this part of the thesis will connect & consolidate previous chapters and lead into the world of data-driven management supported by e.g. data analytics Business Intelligence & simulations.

4.1.2 Challenge: Aviation field in era of globalization

[24] According to IATA (International Air Transport Association) annual report of 2018, who represents 275 airlines comprising 83% of global air traffic, 2018 airline traffic continued to grow strongly at +6.5. The transport performance of European airlines, at +6.6 percent, developed somewhat less dynamically than in the previous year (+8.2 percent), but was robust. The reasons for this are the insolvencies of the years 2017 and 2018 as well as increasing uncertainties in the economic development.

Further the growth of European aviation sector was around +6.2 percent. Despite the political and security problems, worldwide most of the bigger airports were able to increase their passenger volume the pressure especially throughout the summer season increasing day by day due to very tight scheduling and air traffic jam in European skies. In total 2018 there were more than 244 million arriving and departing passengers alone in German airports and this is symptomatic for the global and above all European market. The

global amount of passenger will increase rapidly in upcoming years mainly in emerging markets as Asia (India, China), South America etc.

The strong growth spurt in European aviation is a fact - with nearly 8.2 percent growth expected for 2019, Europe even would passed the Middle Eastern market. In other words nowadays more people than ever before travel by plane. Worldwide, the airlines carried 4.1 billion passengers, as the UN aviation organization ICAO announced last year. That was an all-time-record, which will be broken very soon in upcoming years. It can be said that this development was supported by the strong recovery of the global economy. The above-average growth of the low-cost carriers also contributed to this. In following chapters below a specific example of Lufthansa Group will demonstrate the major data-driven digitization related topics within such airlines.

Global Passengers (billion, segment basis)
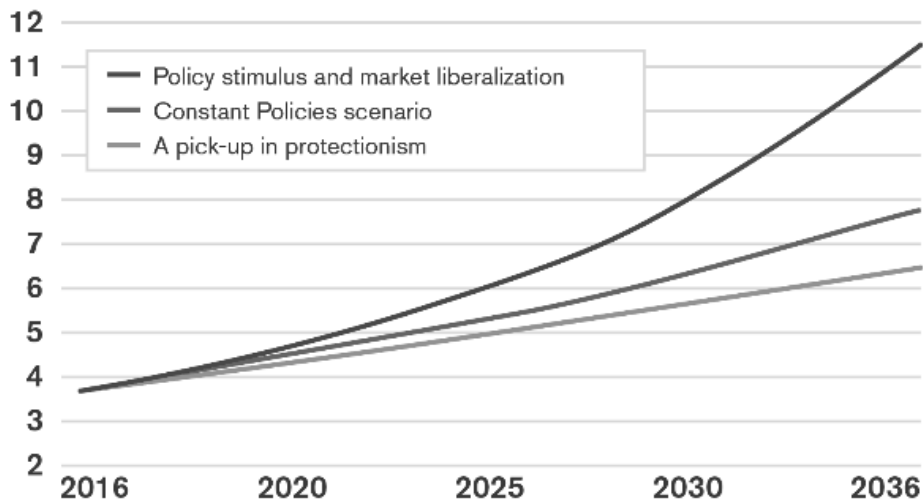
Picture 34 © IATA

4.1.3 Challenge: Development of market

According to ICAO report [26], 2017 around 1.2 billion passengers had flown via low-cost carriers, which is about 30 percent of the total passenger volume worldwide. Legacy airlines as Lufthansa alone recorded an increase in passenger numbers of 18.6 percent to 130 million passengers throughout the Group alone in 2017. Lufthansa as well as its subsidiaries such as Eurowings, Swiss and Austrian Airlines also benefited from the high ticket demand for the competitors.

This might be a positive news for the industry but as the number of passengers and flights increases, the situation in the European aviation industry becomes increasingly critical. [25] Especially 2018 throughout the summer season there has been a significant increase of long delays 6 cancellations compared to previous years. As a result, until 14 May 1466 flights e.g. in Germany had a delay of more than three hours with a lot of hassle & uncertainty for the passengers and crews.

One other bad side effect of the increasing market is that the amount of technical problems on aircrafts increase additionally in summer time because the flight schedules are very tight so that only mandatory maintenance can be undertaken. Due to lack of ground time nearly no preventive maintenance based on predictive maintenance data models and simulations can be done. As if it was not enough a large number of thunderstorms, weather/natural circumstances caused additional delays and cancellations.

Due to lack of data-driven management but also operational working focus it was not possible to react or pre-plan properly on such circumstances. Mainly because beyond cockpit and several other technical areas such as air traffic controlling, security measurements many other key fields in aviation industry is even to date quiet manual. One of those area e.g. would be ground handling, which is one of three major business areas of an airport. Airports offer a wide variety of infrastructure and services, including ground handling. In addition to the entire process of handling passengers and luggage, this also includes the apron handling of the aircraft. These essentially include:

> supplying the machines with fuel and possibly de-icing agents
> the apron control, in which all technical and external conditions of the aircraft are checked
> The handling of a passenger is primarily about the handling of the ticket and hand luggage.
> But also often the allocation of seats and measures in case of delay belong to the organization at an airport

The ground operational processes e.g. are the connecting element between aircraft en-route operations and airport infrastructure. An efficient aircraft

turnaround is an essential component of airline success. It mainly focuses on the aircraft turnaround and passenger processes. Furthermore, key challenges for current aircraft operators, such as airport capacity constraints, schedule disruptions and the increasing cost pressure, are highlighted.
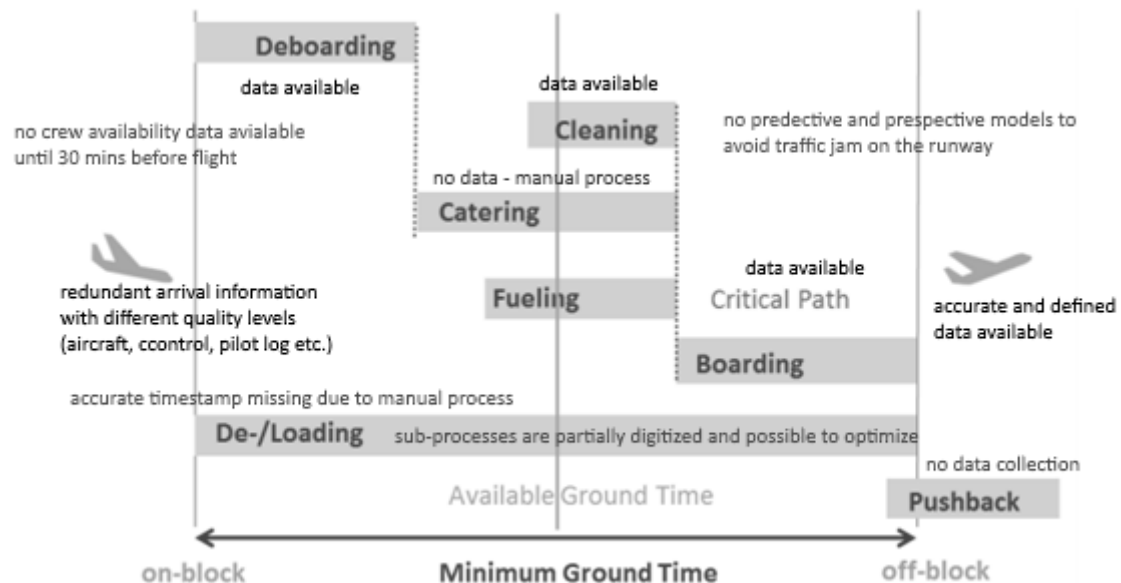
Picture 35 © Lufthansa – modified by author

The modified picture above illustrates the actual situation of many airlines and airport provider worldwide, which is a data collection and quality problem of such important fields as ground handling.

Due to lack of data collection, validation or even alignment about the accuracy, often legacy airlines cannot optimize many influential fields. It pretty much shows as-well that optimization driven low-cost carriers as Ryanair have the same problem since it has one of the biggest amount of delays and cancellations alongside with legacy airlines.

4.1.4 Key factor & success criteria: Data processing & digital strategy

The fact that the aviation sector has been generating massive amount of quantitative data since decades and the majority of cockpit and flight technologies are mainly automatized does not mean, that it was beforehand possible to use those data to improve and adapt business but above all services that are provided.

4.1.4.1 Digitization through predictive maintenance

In order to stay the major airline in European market providing travelling facility to millions of people Lufthansa group e.g. has put major focus on data-driven areas. About two years ago, the company began to simulate digital twins of aircrafts, which is called Aviatar. While the real plane is flying or standing on the ground, its twin forms a huge data packet on a server. The aircraft and Aviatar form a pair that complements each other from production to decommissioning. With modern information technology, the sensors and the processing of large amounts of data make it possible that the maintenance of an aircraft can be planned long before technical issues come up, which helps optimizing flight schedules so that during summer time more time between flights are available. Major focus of Aviatar lies on:

> ❯ What happened, where and when – Descriptive Analytics
> ❯ Why did it happen – Diagnostic or Prescriptive Analytics
> ❯ What is likely to happen – Predictive Analytics Guided actions and steps – Machine Learning, AI and Cognitive Learning

This technology developed in Hamburg exemplifies digitization in aviation. According to experts, it has the potential to change the industry and Lufthansa is not only using it for itself but much more selling it to other competitors as well, which has been a huge market gap until now. Aviatar is a lot about collecting and using data to increase safety and reliability, improve understanding of components and materials, provide more punctuality in air traffic and shorten the life of defective aircraft. Data should bring improvements - for passengers, but especially for businesses. Below a few offerings of Aviatar:

> ❯ Open, Modular, Neutral – Choose your portfolio of solutions
> ❯ Secure and reliable – Customer decides which data will be used and stored by AVIATAR
> ❯ Browser-based and available on tablet, phone or computer
> ❯ Simple and self explaining interface
> ❯ Customer can participate in the developing process
> ❯ Strong increased aircraft availability, less interruptions
> ❯ Reduced expenses (logistics, MRO and asset) by scheduled events instead of unscheduled
> ❯ Increase safety, decrease cost & benefit from a wide field of applications

In collusion it can be said, that solutions as AVIATAR helps airlines to implement new sensors or data collecting points to prevent for unprecedented issues or even to optimize/redesign existing challenges through data

processing and statistical models. It can be seen as the major provider of predictive analytics insights within the Lufthansa Group.
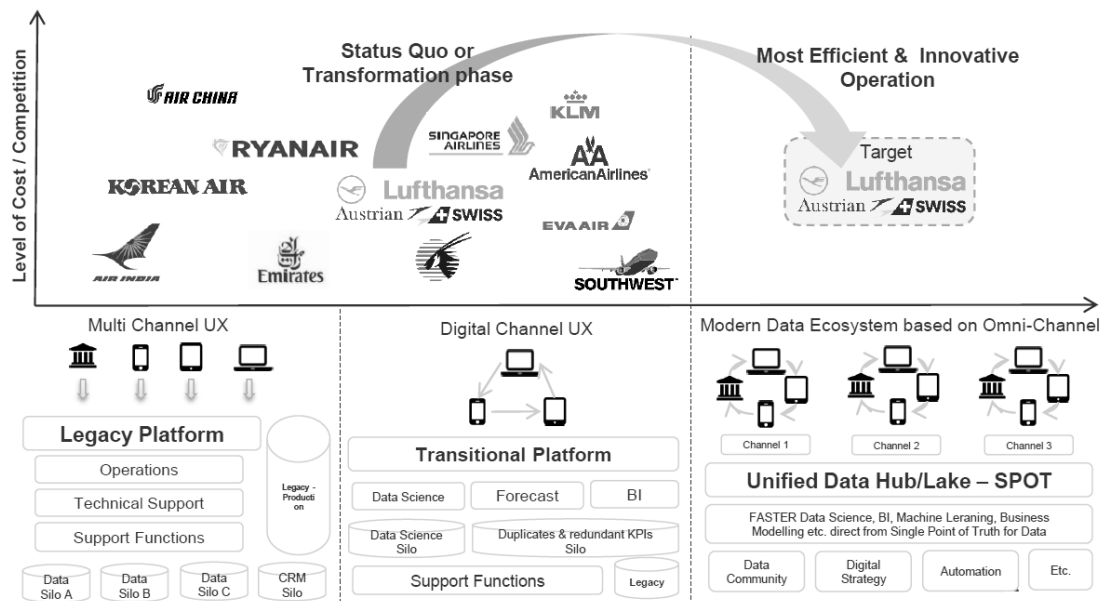
Picture 36 © Alin Kalam

4.1.4.2 BI build the data-driven bridge between operation & management

Business Intelligence (BI) is a set of methodologies, processes, architectures, and technologies that transform and combine raw data derived from the market (external data) with LHG data (internal data) into meaningful and useful information which can be used to enable more effective strategic, tactical, and operational insights and decision-making. Purpose is to provide historical, current and predictive views of business operations as single source of truth (SSOT) for the whole company. Common functions of Business Intelligence technologies include:

> Business performance management (dashboards, benchmarking)
> Reporting (ad-hoc reporting, standardized reporting, monthly reporting, regulatory reporting, online analytical processing
> Advanced analytics (data mining, process mining, text mining, predictive analytics and prescriptive analytics)

For all above many companies use various BI technologies that can handle large amounts of structured and structured data. BI technologies help identify, develop and otherwise create new strategic business opportunities. Aim is to use data and BI technologies in process of identification of new opportunities and implementation of an effective business strategy based on insights that can provide LHG with a competitive market advantage and long-term stability. Exact technologies are described in Technical chapter.

This document will:

- Outline Business Intelligence governance process to carry out successful change and release management
- Define an enterprise/institutional governance council to oversee efforts in managing Business Intelligence across the organization.
- Define roles to govern and own decisions about the Business Intelligence.
- Define technology enablers to support the organization.

Business Intelligence Software makes it easy to aggregate, evaluate and visualize data. The visualization plays a decisive role here. Colors, gradients, graphs and different large shapes make it easier to understand what the data says and what actions are required. However, the use of these tools requires a rethinking of the participants, because many do not want to give up their Excel spreadsheets or PowerPoint presentations, it said in several panel discussions. The new programs process more data than a spreadsheet with Excel allows. Too often needed data would often be in different places (silos) in the group and will not be merged or released. But the larger the amount of data analyzed, the more valid the statements are. Also necessary is a harmonization of the "language" within the group. Different abbreviations, term definitions and key performance indicators (KPI) prevent meaningful evaluations at some points.

Previously mentioned component Data Governance will ensure that data is correct and consistent across the organization. This document will:

- Define an enterprise/institutional governance council to oversee efforts in managing data as an asset across the organization.
- Define roles to govern and own decisions about the data and Business Intelligence.
- Outline governance operating processes to carry out policies and procedures.
- Define technology enablers to support the organization.
- Align to security, data life-cycle management, and quality standards.

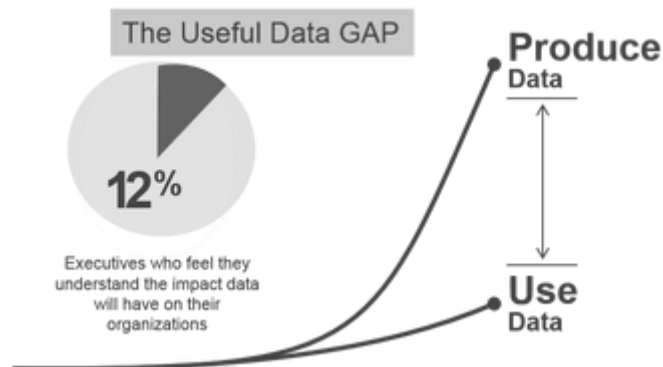**4.1.4.3 Why BI & Data Analytics alone does not solve all problems**

The major major strength of BI is that it helps to visualize facts, figures, findings, results of e.g. predictive data models etc. in an understandable way for management for decision making. Nevertheless also here lies the biggest "weakness" of BI, because for most of the part it cannot interlink between visual driven management and operational streams.

In other words often it does not really help to have solutions like Aviatar in place, when it is not possible for management to correlate their decisions with real-time findings out of data. According to operation manager of Swiss In summer time it would be e.g.:

"Calculation and cancellation of flight X upfront to prevent customer suffering, therefore it is possible to compensate property to keep the customer satisfaction high and as airline it is also possible to concentrate on flight Y which generates more revenue"

A research of practicalanalytics shows that only 12% of management board members across western industries understand the usability and true impact of data on their organization. This the reason why it is very important to build a well-thought holistic embedded platform where data analytics, processes and visualizations can come together.

As legacy airlines often fail to implement such solutions due to traditionalistic organizational structures, it is very likely that their data-driven transformation approaches are very likely to fail or labeled as "less effective" than expected. Seen from this aspect it is mandatory to develop and implement new or best practice approaches such as mentioned below.

4.2 Solution approach: Digital cockpit (embedded BI)

Many current BI projects could in Lufthansa Group can be seen as the market standard. At this stage legacy banks of Central Europe might be on the same level of data-driven digitization as legacy airlines such Lufthansa. It is very symptomatic that similar-sized companies usually implement either Business Intelligence standalone or Data Analytics in many areas of business, mostly decoupled from each other and barely responding to holistic synergy effects of the market & customer. So it was not possible to find out a holistic program, transformation project etc. within the Lufthansa group, which is the main driver for the following solution approach in this thesis. In other words, with the current setup at Lufthansa:

- It is hardly possible to cross-communicate within the group and due to lack of aligned KPIs (financial, operational or marketing)
- It is hardly possible to align on facts figures & insights, which slows down e.g. procurement process etc. with synergy effects

- There is no such single point of truth of data existing, but much more data silos in every airline of the group, who again have many sub-layers of data silos below in most of the business departments mostly not communicating with each other & etc.
- It is nearly impossible to align management boards of each airlines in such data-driven committee to e.g. drive important strategic development of the company. One particular example could be the purchasing process of new Airbus A320 NEO fleet, where an initial order will be placed before even considering the economical & efficiency aspects of the group airlines without a lot data-driven support in decision making

Nonetheless the major challenge could be though, that it is currently also not possible to connect data-driven management to operational activities in required speed. An example could be the "cancellation process" of flights in very heavy-scheduled summer window as mentioned in introduction of this chapter. During the research work, it was possible to **collect data** from a particular airline of Lufthansa Group to demonstrate a possible solution to this problem.

For example there could be a BI & Data Analytics (Embedded BI) based hanger dashboard, which might deliver the most important key performance indicators in one overview for management and operational employees as engineers. No matter if on Laptop, smartphone, tablet or Apple Watch both parties could see:

- Which aircraft are currently AOG (Aircraft on the ground due to technical problems)?
- How many cancellations, delays are owed to a certain aircraft type?
- Engineering KPIs such punctuality, reliasbility, dispatch reliability, cost of failures, active work packages etc. to react and reprioritize tasks when irregular circumstances happen

As mentioned before, solutions such as AVIATAR could build a solid fundament for related innovation. Such initiatives usually should run group wide with the over-all strategic intention to build an aligned layer of valid KPIs rooted in a "single point of truth for related data". In further ongoing this could become the the major driver for the early and systematic identification of deviations in strategically significant investments, projects and the triggering of corrective actions. Below a self-adapted illustration of Aviatar, Digital Cockpit, Data Ecosystem in a nutshell:
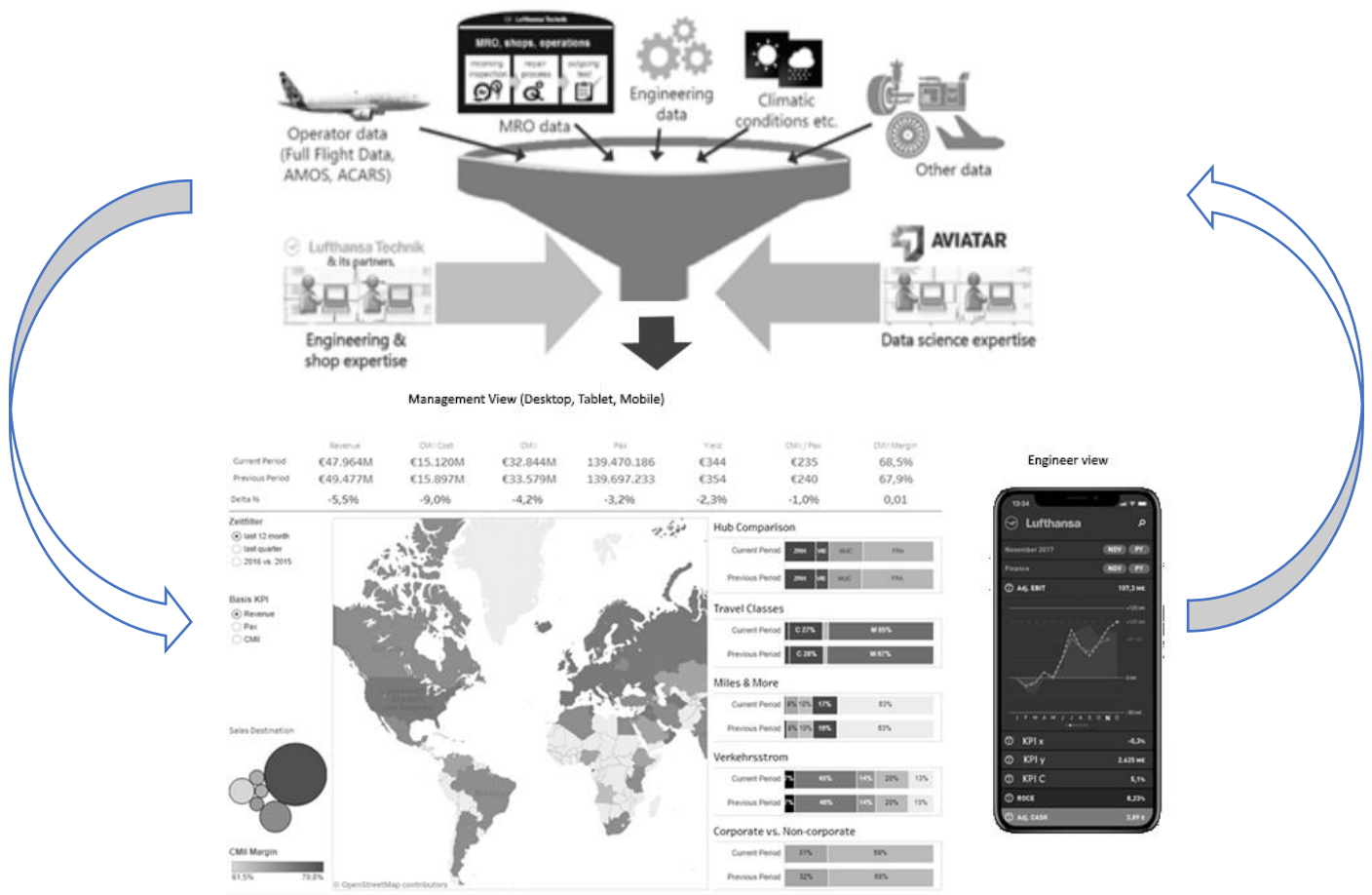
Picture 38 © LH – modified by author

4.2.1 Use Case with data from Airline

In order to create a proper solution backed up by the hypothesis and research questions of this thesis, an airline of Lufthansa Group was requested to provide some test data from daily business. Those data mainly are consisted of basic (non-critical to competitors) information about punctuality, delay, and arrival etc. data, which has been transformed by author (impured with fake measurements) to fulfil DGPR belongings.

In context of scientific solution finding, it was possible to create an initial version of (Embedded BI) platform out of those data, which can be seen as a prototype of a bigger solution. The basic aligned Use Case would be to provide:

- A digital cockpit (Interactive BI dashboard) for Management to leverage from the potential of "data-driven management"
- An associated mobile view for the operational engineers in hanger to create a bridge between management and operational streams

4.2.1.1 Example of provided data:

year,month,day
Date of departure

dep_time,arr_time
Actual departure and arrival times (format HHMM or HMM), local tz.

sched_dep_time,sched_arr_time
Scheduled departure and arrival times (format HHMM or HMM), local tz.

dep_delay,arr_delay
Departure and arrival delays, in minutes. Negative times represent early departures/arrivals.

hour,minute
Time of scheduled departure broken into hour and minutes.

Flight
Flight number

origin,dest
Origin and destination. See airports() for additional metadata.

air_time
Amount of time spent in the air, in minutes

4.2.1.2 Actual data

Arr Ap Actual	Conseq Type	Delay Code	Delay Time	Initial Conseq	Chargeable	Fn Carrier	Fn Number	Delay Type	Occdatetime	Occdate	Distancekey	Categorykey
ZRH	D	43	6	N	N	OS	0567	PSBD	14.03.2017 19:19:00	14.03.2017	10	1
ATH	D	43	17	Y	Y	OS	0801	TI	16.03.2017 09:51:00	16.03.2017	10	1
TXL	D	52	7	N	N	OS	0275	TI	19.06.2017 15:40:00	19.06.2017	10	1
KRR	D	46	45	Y	Y	OS	0607	AC	05.01.2019 09:58:00	05.01.2019	20	2
OTP	D	04A	12	N	N	OS	0791	PSBD	06.01.2019 14:27:00	06.01.2019	10	1
CAI	D	43	3	N	N	OS	0863	TI	09.01.2019 11:27:00	09.01.2019	20	1
CAI	D	46	109	Y	Y	OS	0863	AC	09.01.2019 11:27:00	09.01.2019	20	2
BLL	D	43	16	N	N	OS	2579	IRR	12.01.2019 13:43:00	12.01.2019	10	1
TLV	D	43	14	N	N	OS	0859	IRR	21.03.2019 19:42:00	21.03.2019	20	1
ZRH	D	46	7	N	N	OS	0561	AC	30.11.2018 06:10:00	30.11.2018	10	1
STR	D	43	12	N	N	OS	0177	TI	05.12.2018 06:32:00	05.12.2018	10	1
LCA	D	41	6	N	N	OS	0831	TI	08.12.2018 09:35:00	08.12.2018	20	1
DUS	D	43	35	Y	Y	OS	0157	TI	14.12.2018 20:18:00	14.12.2018	10	1
FRA	D	43	12	N	Y	OS	0121	TI	20.12.2018 06:46:00	20.12.2018	10	1
ZRH	D	43	8	N	N	OS	0561	PSBD	19.01.2019 06:20:00	19.01.2019	10	1
FRA	D	46	5	N	N	OS	0125	AC	27.01.2019 16:40:00	27.01.2019	10	1
DNK	D	43	8	N	N	OS	0675	MTE	01.02.2019 09:36:00	01.02.2019	10	1
CDG	D	43	6	N	N	OS	0417	IRR	04.02.2019 16:33:00	04.02.2019	10	1
FRA	D	05C	21	N	N	OS	0133	LP	26.02.2019 07:31:00	26.02.2019	10	1
CPH	D	51	54	N	N	OS	0307	PSBD	20.03.2019 20:16:00	20.03.2019	10	2
VIE	D	41	983	N	N	OS	9104	MNSS	25.09.2018 10:45:00	25.09.2018	10	3
DUS	C	41	0	Y	Y	OS	0155	CCT	25.09.2018 15:40:00	25.09.2018	10	1

Picture 39 © Alin Kalam

4.2.1.3 Example of data preparation with python notebook

Data preparation is done using the Jupyter Notebook interface using programming language R. To help make this more concrete and indicate what is possible, here are some examples of showing Jupyter Notebooks being used for data preparation using following function:

```
library(XLConnect)
library(xlsx)

The procedure was taking data from an Excel file with a command like this:

dati <- read.xlsx(file, sheetIndex=1)

Then, with downloadHandler(), the App should provide the final results in .xlsx extensions
too. These lines should complete the task:

output$downloadData <- downloadHandler(

  filename = function() { paste0(Sys.Date(), '.xlsx') },
  content = function(file){
    fname <- paste0(file,Sys.Date(),".xlsx")
    wb <- loadWorkbook(fname,create=TRUE)
    createSheet(wb, name = "Sheet1")
    writeWorksheet(wb,passData(), sheet = "Sheet1")
    saveWorkbook(wb)
    file.rename(fname,file)
  }
)
```

4.2.2 Challenge statement

According to several interviews, the airline very common problems as some of them mentioned in in chapter 4.1. Throughout summer times it was not possible to provide to management and operational personal the same view of:

- AOG – Aircraft on Ground because of technical, weather and etc. issues
- Availability – Visible availability of Aircrafts beyond utilization between flights
- Predictive maintenance – E.g. a predictions about clearance of aircraft while trouble shooting of technical problems begin

Further as result of scientific investigation, it could be found out that, beside technical infrastructure and challenges of data handing a known issue has been existing since years, which is the lack of KPI definition. The company has been struggling for years to create a "single point of truth" of data to prevent dualities or parallel definitions of KPIs in different departments. Several approaches to build an ecosystem of "single point of truth" for data has been running since couple of years, which supported the hypothesis of this thesis.

Therefore as major challenge also the definition of KPIs was taken into note. To create an initial version of interactive dashboard for management & engineers, the software TABLEAU Version 12 was used, which offers a very easy handling of prepared data into visual graphics. In picture below a demonstration of nearly automated handing of data and creation of dashboard only using the prepared data.

Picture 40 © Alin Kalam

4.2.3 KPI definitions & solution:

Key Performance indicator	Aircraft on the Ground (AOG managing time)
Definition	Shows the portion of the AOG which in one have been resolved over a certain period of time. Formula: $$AOG_{managing\ Time} = \frac{AOG\ solved\ within\ category\ time}{Total\ number\ of\ AOG}$$ It should not only provide a holistic view of AOG times for management, but also track the current rate of "compensation". (If a flight delays more than 3 hours, the passenger must be compensated within a time frame of 3 years.
Strategic Goal	By tracking this KPI, it will be possible for e.g. financial department to track the potential costs for C - compensation in upcoming budgets, since not every customer insist on a compensation. But if a flight is delayed more than 3 hours, it must be cancelled and the aircraft is set on "AOG" modus. Other categories as B > 45min should express, the uncritical AOG frames, which e.g. does not cost anything for the airline because they are usually caused by airport handling, weather etc. circumstances Category A<=45 would be the midrange between B & C which has slightly financial effects.
Solution Approach with Tableau	 Picture 41 © Alin Kalam

Key Performance indicator	Return to Service
Definition	Shows the accuracy of the statements regarding the return to service time in AOG case.$$\frac{\sum Events\ per\ Category}{\sum Total\ Events}$$It should provide a prediction of clearance in case of technical issues on aircraft (AOG reason: technical)
Strategic Goal	This KPI should be used by flight planning unit, management & engineers so that all parties have the same view to re-scope, re-plan etc. in case of an AOGExample: During summer time there is a very high utilization of aircrafts and there are more technical problems that usual. In such case, management and flight planning unit have to pro-actively predict the releasing time of aircrafts after problems are solved in upfront. Delivered predictive measurements and categories from AVIATAR (chapter 4.1.4.1) were used in this KPI:Multiple Next Info, No Info, Ready in Time, Late, Early, Multiple Advise Times, No Advise Time
Solution Approach with Tableau	Picture 42 © Alin Kalam

Key Performance indicator	Event Tracking
Definition	Shows the proportion of time in which one Airplane is in operation, located in Maintenance is located, AOG, by a GTE is blocked or is available beyond. $$\frac{\sum Time\ per\ Categorie}{\sum Total\ Time_{Day}} \times 100$$ $$Total\ Time_{Day} = AC_{\sum} \times 24h$$
Strategic Goal	This KPI should provide the overall availability of all fleets utilized by the airline for management and current events as "Technical Problem: aircraft XYZ is on ground due to XYZ problem". The reason behind is less management driven but much more operational, since engineers have to know immediately when and where a critical issue arises. Therefore it has a less BI related and more real-time event tracking requirements connecting management and engineering as a bridge. The simple goal is to track all technical problems and find solutions in real time as ordering spare parts or allocating manpower to several tasks.
Solution Approach with Tableau	Picture 43 © Alin Kalam

Management View

A demonstration of management view created in Tableau. Others than "AOG managing time" & "Return to Service" are pseudo measurements used to illustrate and support the optics. (Picture 44 © LH – modified by author)

Regularity
98,12

⊘ Cancellations
389

€ IrregCosts
4.281.750

Technical Performance CW24:

In der Woche CW 24 gab es einige AOG's auf A320, E95 und DH8 Flotte. Die größten Auswirkungen gab es durch eine fehlerhafte Modular Avionics Unit auf der LWP in GRZ und Probleme mit der „Landing Gear Control Interface Unit" auf der LDA wo nach aufwendigem Troubleshooting ein defektes Safety Valve als Ursache gefunden wurde.

Aircraft Availability

Legend: AOG | GTE | Maint | Ops | Available

ER19 — 13.9%, 44.6%, 39.7%
D84 — 4%, 17.9%, 42.6%, 34.8%
A32F — 18%, 46.8%, 33.2%
777 — 17.8%, 70%, 11.5%
767 — 9.2%, 82%, 7.7%

Available D84: 34.8 %

AOG managing time

A <= 1h 4.46%
B - from 1h to Compensation 11.16%
C - Compensation 84.38%

Return to service quota

In Time 0%
Early / Late 49.5%
Others 50.5%

test.now-app.io/

Engineering View

A demonstration of engineering view created in Tableau. It simply shows the real-time events delivered by AVIATAR and based on KPIs on management view and could be seen as the operational wing of (Embedded BI) providing the right tailored view relevant to each group of users. (Picture 45 © LH – modified by author)

4.3 Possible impacts

It is needless to mention that the impacts of "Digital Cockpit" (including management & engineering view) initiative could be a key game changer to revolutionize the culture of working and decision making between management board (incl. middle & lower management) but also operational streams, engineers, crew etc. It could not only driving fast data-driven decision making and operational work but much more it has become the major factor in driving digitization (process automation, optimization, data collection etc.) itself and bringing improvements in nearly every corner of the company.

Suddenly data quality, governance and over-all interpretation will get attract a proper attention, which might not be the case currently. Such holistic approach proves, that data-driven digitization is not only a matter of cost saving or ruthless investing but way more than that. In case of a legacy airline as Lufthansa Group it is an intelligent combination of "investing at right places" such as building an appropriate future target architecture (ecosystem") based on hiring data specialists, Implementing Data Governance & management methodic within the company and providing those knowledge based & data-driven insights to a broader embedded environment for consumers of many domains. According to market standards and my own experience of nearly a decade such data ecosystem could be consisted of:

> Data Lake – for building data mart, the so called "single point of truth" (SPOT)
> Data Processing – Data Mining, Data Science & AI to create e.g. predictive insights
> Process Mining – Tools and methodologies to optimize/design processes
> BI – Business Intelligence Software solutions to visualize KPIs etc. based on SPOT
> Embedded – Applications to combine Data analytics with BI and enabling interaction in both directions

According to several interviews with middle management of Lufthansa, in many parts of the group currently the concept of agile working methods is being tested. Especially the IT and software development domains of the group have been implementing agile methods to their work, which makes Lufthansa Group one of many hybrid companies using agile and traditional approaches in data-driven digitization pipelines. This is also the standpoint of many other big companies, which says:

"A healthy combination of agility and legacy methods connected with data-driven ecosystem approach will be the key of digital transformation.

Additionally the so called data-driven management/decision making will round up any surely accelerate this process".

4.4 Summary

In previous pages it is illustrated how a potential Embedded BI Dashboard for management and engineering could look like. As mentioned beforehand, creating this bridge between business areas has been a major challenge for such legacy companies, which is the main focus of the solution approach.

Nevertheless often it is not enough to have a nearly perfect ecosystem and digital strategy in place when companies do not manage to take regard of ethical standpoints of data-driven business transformation. Next chapter will deals with the topic of ethical, moral aspects of such transformation approach, which must not be neglected due to enormous attention of legal bodies, public and media.

CHAPTER 5.0

5.1 RESEARCH AREA 3 - Ethical aspects of data-driven digitization

Hypothesis 3: Ethical Aspects of data-driven transformation when neglected build serious threat to existing business and can be destructive for future development.

Historically seen there has been a lot of ups and downs of the technology industry and ever since change has been the only stable constant. As our economic structures are made from- and for human, it is obvious that any kind of dehumanizing (e.g. automation/digitization) has had major effect on us in previous times. Many would admit that the current hype of *"digitization" is actually nothing new- but much more an intensification of ongoing automation through data & information driven technologies.* There are several processes that together can be seen as the starting point of digitization. This includes on the one hand, the development of the microchip, which, according to Moore's Law, doubles its computing power every one to two years. With it ever more powerful computers, robots and sensors can be created. The specific starting point of data-driven revolution might be the emergence of the internet in 90s.

In the world of industry-based digital technologies and the related automation in past decades a major challenge has been answered with the exclusion of so called Industry 4.0, which again is in many ways same domain only connected with new revolutionary production concepts as 3D printers etc. In the area of end users & economy, the rising digital technologies mean the removal of manual work, faster data gathering/processing, gaining knowledge

out of Information that has roots and validity through data & borderless exchange of it. In conclusion data-driven digitization enables:

- New scientific insights to tackle the global challenges (climate change, pollution, hunger)
- Improvement of information and education for all and incising of transparency in politics, administration and economy
- Qualitative improvement and acceleration of decisions
- Optimization & increasing effectivity of production sites
- Creating new high earning jobs
- Strengthening the science and business location

But not every aspect of digitization is positive and there might be some ethical but as well moral standpoints of ongoing data-driven automation and e.g. dehumanization of business [29]. Except of sweeping jobs away, the rise of data-driven hype has been a door opener [30] to e.g. cybercrime & increasing spying, intervening of privacy and etc. with a huge conflict potential across borders. Some are mentioned below:

- Depersonalization and loss of social exchange
- Informational complexity hinders democratic control promoting e.g. spreading of fake news
- Vulnerability of information technology systems of companies and countries
- Digital decisions / interventions are no longer comprehensible
  Due to the complexity and quantitative amount of data
- Misusage of technologies and data processing methods (e.g. people analytics to spy on employees)

In upcoming chapter the most important ethical & moral aspects of data-driven digitization will be examined. Few cases will show that e.g. people analytics can impact companies in very positive ways if the moral barriers are kept and organizations set up their framework properly, which also might apply the other way around. As the other side of the coin a specific case from industry will show, how destructive can data-driven digitization ca become if moral and ethical barriers are neglected.

5.2 Ethical/moral & legal/regulatory baselines

On May 25, 2018, EU-wide General Data Protection Regulation GDPR was activated. The basic intention was to provide every EU citizen more control over their own data. This affects companies of all sizes, from sole proprietors to corporations. In addition, the regulation applies to all companies processing data of EU citizens, including those based outside the EU. In a matter of fact

it does set some high standards in protecting consumer interests but at the same time not a lot ethical/moral barriers beyond law is implemented. So what are the ethical baselines for data-driven digital transformation? In an interview with Chief Security Officer of BAWAG Group said:

"Data & information security is the technical baseline in ensuring the confidentiality, integrity and availability of right information and privacy of customers. From our perspective the highest level of preventive safety measures can protect data from e.g. unauthorized access or manipulation and save lot of time and money that should be invested in innovation, customer & employee satisfaction"

Further, according BAWAG CISO, GDPR is a worldwide unique framework that ensures the highest standards and protection needs of data & information of customers, employees & private persons by law. It sets at the same time moral and ethical borders of collecting and processing data for certain innovations. To understand the complexity of data-driven transformation of companies and synergy effects with ethical/moral standpoints, firstly it is important to analyze the example of worldwide job sector, which has been massively effected by automation throughout last decades, because over the course of time not only jobs conditions usually has been affected but such transformation has had major impacts on living circumstances of people.

Share of Jobs with High Automatibility (%)
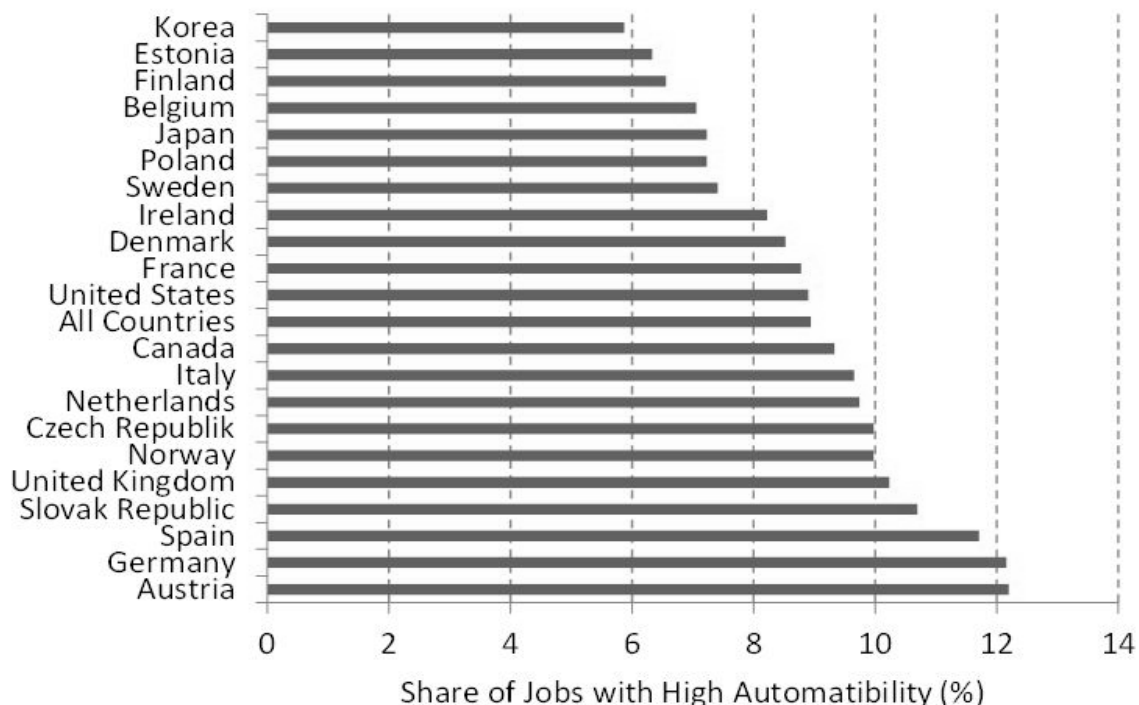
Source: Arntz, Melanie, Terry Gregory und Ulrich Zierahn (2016), The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis, OECD Social, Employment and Migration Working Papers No. 189, Paris.

Picture 46 © OECD

Many people fear that computers and robots will eventually make their jobs obsolete. A new study by the Organization for Economic Co-operation and Development (OECD) now suggests that this concern is not unfounded: about half of all jobs in the 32 countries that participated in the investigation are threatened in whole or in part by data-driven algorithms and machines. According to OECD about 14% of the jobs examined are classified as "highly automatable". This means that the likelihood that the work will be taken over by machines or programs in the future is more than 7%. This high risk affects more than 66 million jobs across the OECD countries and the ongoing digitization will only expand and deepen in upcoming decade. This facts clearly proves the ethical dilemma between innovations that improves peoples da-to-day lives but on the other hand destroying huge amount of jobs, which will push many into poverty. In case of elderly citizens of society often it is not easy to retrain and learn new methodic & profession. Also it does not need any high level of statistics to underline the fact, that there will be certain amount of citizens never getting higher educations, who might kept trapped into the world of unemployment, because digitization will make their jobs obsolete.

As a revenue generating company it is often hard to correlate with needs of the common employee and his fears of losing job. Often is automation less of innovation and customer orientation but only ruthless unethical cost cutting agenda. As mentioned before in previous chapters, those who does not understand the fact that digitization is everything but only optimization and cost cutting or spying customer & employees, will most likely lose huge amount of market shares in upcoming years either by paying huge amount of penalties or by losing customer loyalty. If a company is not caring about reshaping & designing the business (e.g. omni-channel approach or adapt the customer journey/experience throughout the whole business), chances are very high that it will be experiencing a rude awakening lately by losing customers and business. In other words very often moral and ethical standpoints can save companies from ruining it´s identity and customer orientation, which can become a financial disaster as well!

In case of big multinational companies throughout developed and emerging markets, it is obvious that often this fact is being left aside. Being ethical and responsible can also mean a higher customer satisfaction (economy made for human from human). The open secret of successful business growth and sustainability might be the healthy and holistic combination of dehumanization of needed areas but therefore creating new opportunities to retrain loyal employees, providing good service to the company for years. A particular example of legacy banks from American banking sector can be mentioned as the light house example on global context why it is a bad business idea not to set ethical & moral barriers while proceeding with digitization.

For the financial industry, the age-old programming language COBOL had in previous decades a great importance. Every day, transactions worth an estimated dollar are processed through COBOL systems, which is often the mainframe/ core banking system of big legacy banks in USA. This involves checking accounts, card networks, ATMs and the processing of real estate loans. With banks aggressively banking on digitizing their business, COBOL was set to be phased out and many elder COBOL specialists either early retired or set into unemployment due to huge amount of costs to keep them.

According to a CNBC report [31] by time with emerging of newer faster technologies, much more focus has been lying on decommissioning legacy systems or build technical infrastructure from scratch (cloud etc.). But although there has been other technologies and more modern languages, COBOL has stayed an indispensable part of big banks, corporations and parts of the US government. Because the powerful computer systems of related companies and authorities were often built in the 70s or 80s and were never completely replaced. The main reason would have been the concept of "Never change a running system", which is even today the highest mantra in IT. As a result many big banks are now doomed to play a much higher price for COBOL specialists on consulting basis and it is a fact that the gap between supply and demand is huge.

5.2 Data-driven digitization done right – people analytics driving meritocracy

Beside customer analytics, the topic employee analytics has been a huge topic of interest. The collection and analysing the behaviour or characteristics of employees on the basis of data to improve company's performance can be generally called as people analytics. After all, companies are only successful if they exploit the potential of their employees, design processes efficiently and react to problems at an early stage. Google was one of the first companies to start using people analytics back in 2009 to drive its data-driven decision-making processes, product development, marketing and pricing. This kind of systematic evaluation of data on employee satisfaction has found its way in many companies

THE 7 PILLARS OF PEOPLE ANALYTICS

① WORKFORCE PLANNING ANALYTICS
② TALENT SOURCING ANALYTICS
③ TALENT ACQUISITION / HIRING ANALYTICS
④ ONBOARDING & ENGAGEMENT ANALYTICS
⑤ EMPLOYEE LIFETIME VALUE & PERFORMANCE MANAGEMENT ANALYTICS
⑥ TALENT ATTRITION & RETENTION ANALYTICS
⑦ EMPLOYEE WELLNESS, HEALTH & SAFETY ANALYTICS

THE FRAMEWORK TO LAUNCH SUCCESSFUL PEOPLE ANALYTICS

Picture 47 © ere.net

throughout the current decade. The positive opinions people analytics has

been great and with the right tools and implementation that respect privacy and ethical standpoints, People Analytics can significantly increase employee satisfaction and retention. While organizational network analysis (ONA) has been used for the past few decades by researchers, consultants and practitioners to understand how people and groups interact with each other, this data new driven approach based e.g. real-time analytics has been one of the most innovative fields for couple of years. According to Ben Waber's book *People Analytics [32]* it is quite hard to connect social data from company badges, email networks, and other sources to understand and learn from employee behaviours. The author uses a series of case studies, many of which are based on his own primary research with social badges, to illustrate several points.

For instance it could be a huge help especially for HR departments around the world to find out:
> Which employee fits in which team?
> Were to install certain office facilities (furniture, meeting rooms, tables etc.) and what are the impacts?
> Which type of communication supports productivity
> What promotes creativity and solution orientation?
> Anonymized employee satisfaction

Beyond HR there is a huge potential of people analytics in several other fields of economy & modern industry and people analytics is maybe the biggest driver of meritocracy in modern companies.

For instance an analytics dashboard which displays the skills, performance and over all competence of an engineer over the entire career will allow his organization to become a true meritocracy. Only in such system the good engineers will most likely get promoted because they're technically better and effective. This is the kind of company which will attract even more well-performing engineers, because they can see that their career progression can be meteoric if they are skilful. A company which is able to measure technical skills properly, and reward them well, has an unfair competitive edge in the market because it will quickly become the employer of choice.

Combined with analytics this type dashboards could be used to prove the value of the training provided by HR department. Every year big companies invest millions in further trainings and education of employees, but studies of eNyota learning e.g. show that the impact is enormous when it is done right. On the opposite it is also well-known that often

COST TO VALUE OF AN EMPLOYEE

key performers performance decrease due to wrong or unadjusted trainings, courses etc. In reality many companies even can't answer basic questions about the training they offer: "Is it better to do the training online or onsite training is the better option?" "Should it be done informally, one on one?" & etc.
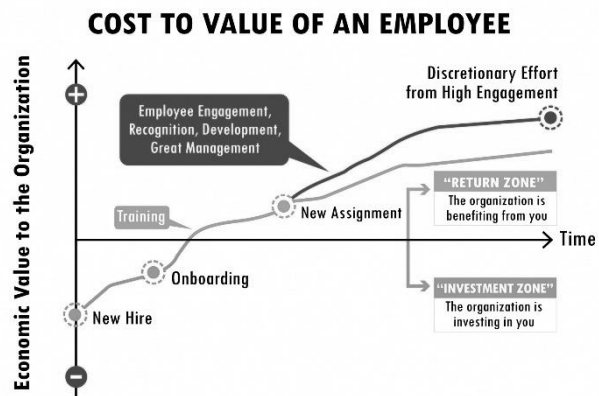
In other words people analytics is not only surveying but also tracking and analysing several standpoints directly connected over-all performance of the company. A specific case from Bank of America shows how people analytics helped the bank to improve it´s 60 years older concept of running call centers, where people are usually trained the same and had relatively similar call qualifications. People analytics should have [33] helped them to find out why nonetheless some call centers had significant better performance ration than others. Based on collected data all of a sudden questions like "what behaviours raise productivity & efficiency?

It turned out that the productivity is directly connected to how employees communicate to each other rather than when or what. According to the result of the project, it turned out that 80% of employees' interaction happened in the 15 minutes when peoples' lunch breaks would occasionally overlap. Bank of America used that information to A/B test a new break system, which was one of the major factor in raising productivity in other low performing branches.

5.2.1 How people analytics is changing working culture

For a long time performance-, effectivity-, efficiency & etc. tracking based on analytics has been reserved e.g. for top athletes as Lionel Messi and co, but soon it should be available to every office employee. The company Humanyze from Boston is has been a success story creating innovative solutions in this specific field of innovation. Instead of just reading e-mails and matching calendar entries, Humanyze wants to know exactly what employees doing throughout their working day. The five computer researchers at MIT Human Dynamics laboratory and the Aalto University in Finland, who founded the

company in 2011, have developed a device that is about the size of a pack of cigarettes and is worn around the neck on a neck strap.

Humanyze has puts several sensors within the device, including two microphones, Bluetooth and infrared receiver and a motion sensor. This allows to measure the volume of conversations, record speech and gestures, and even determine how close two conversationalists are. The content of the calls does not record the box, but the microphones are set extra. [33] Humanyze is not so much about what someone says, but about the way it is said, because often "communication makes up a large part in many professions and is therefore also a key criterion for success and employee satisfaction," says Humanyze CEO Ben Waber.

It worth to mention though that the device is not recording the conversations between employees, much more it delivers a real-time audio processing based on Big Data to find out what is going on in the company and which topics are currently relevant. The decide includes also movement analytics, so more information about
Their physical activities. It is also important to mention that the company does not provide any individual metrics to its customers (companies). Individuals own rights on their personal data and such analytics is done on opt-in basis. However it is possible to combine those metrics with KPIs from companies, so they find out:

- What specific groups/segmentations, e.g. best performer of the company are currently doing?
- How much of cross-department communication is currently being done, that have already an existing communication gap? (e.g. engineering & marketing)

According to Humanyze, companies and individuals get feedback on this behaviour and within their technical environment the data security is guaranteed. Their business is strongly built on analytics capabilities and providing insights rather than selling hardware to customers.

5.3 Impacts when unethical data-driven business transformation – Austrian Post

Late 2018 it was explored by research platform "Addendum" that Austrian Post has been trading extensively with data from its customers. Due its market monopoly and Austrian Post has been saving a huge amount of data for decades. Addendum proved that the Austrian Post has been selling predictive data of its customer to political parties throughout many previous elections (predictive customer profiling about the political attitudes, political interests based on customer data/buying and ordering behaviours, but also even more

sensitive personal data that have been generated on top of existing data sets). The Post Austria claims to manage around three million records and profiles. According to Addendum it has created profiles of political affinity about 2.2 million Austrians exclusively for previous elections.

A detailed investigation of Austrian Law authority, it became clear that the Austrian Post has been selling this data since years, which is one specific product of their digital business. According to the data protection authority for certified companies as Austrian Post it is principally not prohibited to sell data to 3rd parties. Nearly every company does nowadays customer profiling, segmentations e.g. to offer adjusted & personalized products/services to their customer, which is not against GDPR as long the customer has granted an opt-in permission. However the authority has also disclosed that there has been PII (personal identifiable data) data such as surname and name incl. address without an approval of customers. Also there has been no information provided to customers about the profiling regardless profiling and selling data about political preferences, which is illegal anyway.

Currently there are further investigations against the Post running and obviously there could be significant penalties because the data storage, processing & political profiling is covered neither by the trade regulation nor by the General Data Protection Regulation. [34] According to chapter 9 (1) GDPR, the processing of personal data resulting in political opinions is prohibited. Art. 9 para. 2 lit. a GDPR requires explicit consent in this regard. It is interesting how Post Austria intends to provide proof of the consent of the data subjects required under Art. 7 (1) GDPR.

5.3.1 Moral/ethical aspects & financial impacts of Post Austria data scandal

It might be not 100% final because no decision of the authority has been made yet, but it is already proven that the Post Austria has been collaborating with major technology providers, research companies to build this specific product mainly for political parties so that millions of customers could be contacted based on personal and predictive likelihood. As it is illustrated in picture, the company has built a new licenced business segment around data consultancy, analytics services and data trading. As reported in various media beginning of 2019, the Austrian data protection authority has initiated a review of the postal services.

It might be common practice to calculate probabilities so companies can send targeted offers to its customers, but nevertheless there should be an ethical barrier to sell it to 3rd companies since it is a touchy issue of privacy. Especially when it is directly connected to political, religious or other private fields of individuals. The potential & impact of misuse is given, which could become a major root cause of systematic disadvantage for many people. As a result not only customers but also investors can turn away their back, which has been the case since the scandal was revealed in November/December 2018. After years of flourishing business and customer gain, Post Austrian lost a lot of market during the very busy Christmas season followed by customer & employee complaints, law suits etc. caused by the scandal. That fact was also visible at the stock market.

Picture 49 © Finanzen.net

This specific incident demonstrates a very weakness of data-driven approach in many companies, because often they invest in newer technologies to extend or optimize their business without taking regard of ethical standpoints. Previous annual reports e.g. (2016) [35] prove massive investments and focus on Big Data, analytics platforms etc. to build new business segments. Researches have shown that the new digital business segment was not coordinated with the ethical/moral codex needed e.g. for customer loyalty. In a matter of fact it was quite the opposite! Except of planning and benchmarking useful services & digital products the Austrian Post invested money in Big Data hype at the pick of its market popularity just to look innovation and digitization driven.

In times emerging technologies data-driven approaches as people analytics, customer profiling might often look profitable and reasonable at first sight, but such case show the serious need of alignment between innovation, profitability and ethics, because otherwise entire company environment, started from customer, employee to public relations and investors, could be affected.

5.3.2 Balancing the risks, rewards of people analytics/customer analytics

Data, algorithms, artificial intelligence - all these are becoming increasingly important in all areas of the economy and society. According to Laurence Collins, a director at professional services firm Deloitte, the ethics around customer & employee-related data is the most keenly debated area with his clients.

Mainly because often it is not clear to all participants what is the true meaning and value of it. While providers offer analytical tools to their customer companies, emphasizing that the software will help the company to change, it is unimaginable for many employees those companies to be a part of this innovation, because in case of a data leak or misuse the personalized insights is threatening from every aspect.

A specific challenge associated with personal data analysis is the loss of individuality. Usually people and their way of working is an extremely complex phenomena. Different employees have different work styles and their productivity can vary from job to job. In this case, data analysis does not really make a difference between alternative working styles or productivity forms. While results may come out, they do not necessarily have to contain information that really helps the supervisor or the company. Another aspect is, if employees are aware that they are monitored and ultimately measured via analytics, it can lead to strange or unpredictable behavior in the data set. For example, some employees may try to play the system by misbehaving, misusing etc. It certainly would create negative incentives for innovation or leads to unconventional reactions of authority. Overall, if employees become

accustomed to the tool, this can lead to predefined behaviors that only aim to be considered negative by the analysis tool.

One of the main objections to the use of people analytics is certainly data protection reasons and privacy needs of employees. In comparison with USA or Asian market, data protection has a higher value in Europe. Especially collection and processing of personal data is very sensitive due to lack of IT security in such companies. Potential data leaks also could promote competitors, who would have a head start through such leaks by seeing major performance related variables within the organization. The risk that personal data is visible for non-related users in plaintext or misuse is of course inevitable and according to the GDPR may cost 20,000,000, or (if higher) 4% [34]of the annual turnover, which builds a huge financial & regulatory risk for companies using personal data regardless if it is PII data. There is a compelling need to carefully manage PII data, both from a compliance and business perspective. There are several factors in working through the PII implementation:

- Not only rely on 3rd party security measurements
- The data privacy / compliance requirement that no users (e.g. data scientists) should not be able to see PII data unless it´s necessary for their job
- Compliance requires that a knowledge Worker (or other member of the business team) working for a tenant (project etc.) can only see data for that tenant.
- In some cases PII reference data needs to be provided to down-stream systems (e.g. along with the results of models)
- In some cases there is value in processing PII data to provide derived (non-PII) fields that have analytic value (e.g. in determining the country of origin of a person's email account)
- Systematic detestation, masking or obfuscation of PII data, that is not needed for analytics insights
- Alignment with legal department & also union to make sure a proper framework for measurements

As mentioned in previous chapters, components as Data Governance & holistic ecosystem based on proper IT security concepts be the perfect environment for such data-driven approach. A few important points to follow would be:

- Data specification and modelling, processing, database maintenance and security, to ensure that data will be fit for purpose and held securely in their secured databases
- Ongoing data audit, to monitor the use and continued effectiveness of the data

> Archiving and final destruction to ensure that data are archived and maintained effectively until they are no longer needed or are uneconomical to retain
> Additional IT security measurements in case of saving data in the cloud
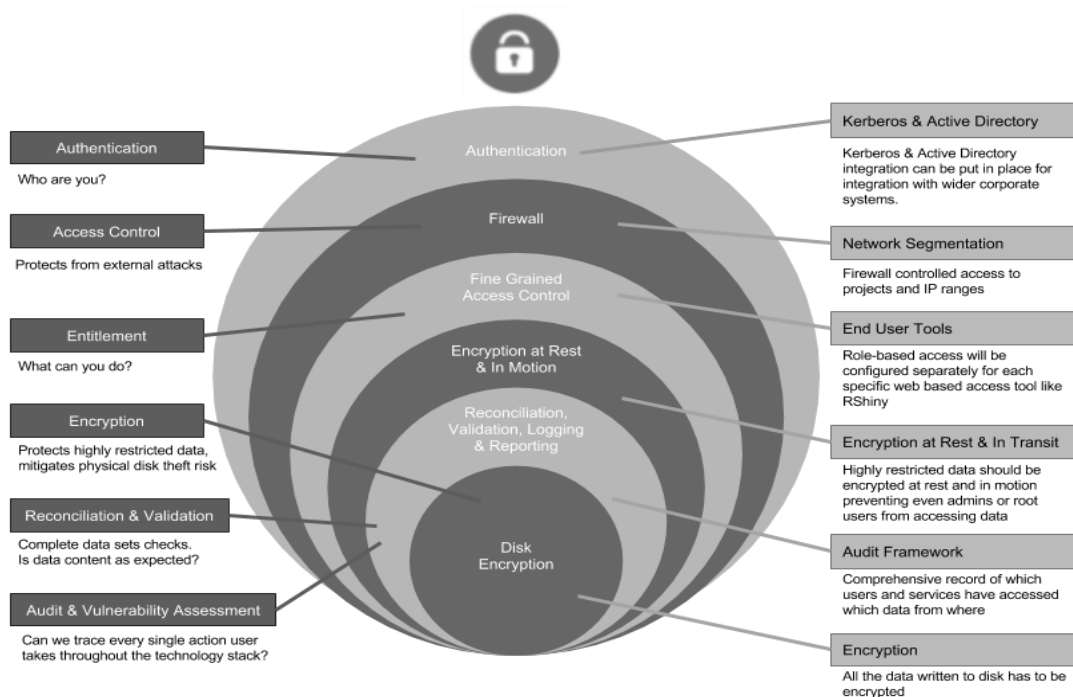
Picture 50 © Teradata

A beyond state of the art It security model covers end-to-end security of the Hadoop/Spark environment, including the operating system, platform software and authentication mechanisms, unauthorized access, and exposures to network attack and data leakage. This topic is especially important when providing processing capabilities on massive amounts of sensitive structured and unstructured data.

> Integrate with corporate Identity Management and Directory
> Support effective security governance processes
> Provide administration and logging facilities
> Authenticate users at Operating System and Application Layers
> Minimise surface area
> Monitor for intrusion
> Secure cluster build out procedures
> Damage and loss prevention

However at any cost the access and the privacy of the data must be respected in mutual understanding and be managed transparently. Data must be saved and processed ethically correct and without discrimination for bigger purposes effecting the company.

CHAPTER 6.0

6.1 Review of Research Objectives

From scientific point of view this thesis has achieved set goals by using several methods to prove major hypothesis right as:

Digitization and Data-driven have very strong synergy effects and in most of cases even depended on each other. In a matter of fact this is the major difference between digitization and automation. While automation has been ever since process optimization and cost reduction driven, digitization is much more customer, emotion, service, innovation driven and has data-driven approach in its core.

Chosen market analysis of India has shown that such data-driven digitization approach can not only change companies but even countries with huge population by effecting daily life of millions of citizens. The detailed research work done on Central European banking sector proved once more, that digitalization is a lot about doing business as a tech company who understands to combine automated, optimized business/technical processes & methods with omni-channel to create sophisticated emotion based customer experience with added value. It is everything but cost-saving agenda only and only those who understood this very message to transform their business and product/service portfolio, will survive the current phase. By naming and research work done on newer technologies, methods etc. it was possible to create a solution approach of implementing Data Governance within an Austrian bank, which can be seen as a prove of concept or show case attempt.

Further 2nd hypothesis was proven right , because chosen methods have shown that impacts of data-driven & ecosystem-based digitization on management, strategy & organization can be very profounding. Through detailed and practical research on European aviation field and the particular case of Lufthansa Group have shown us the meaning and positive impacts of having a data-driven management board as the market situation in aviation field has been critical in recent years due to competition but also surplus in the market. By answering related questions, the major objective of this area was achieved. The solution approach of Embedded BI was backed by industry data from an airline was created for this thesis only. It demonstrates how a data-driven bridge can be built between a data-literate management board and operational streams of the company. By creating a BI Dashboard for management and mobile application view for engineers, an embedded BI solution was invented, which can be seen as a prototype providing important KPIs with right view and details for both areas of the company.

Last but not least by reviewing current market situation and scientific research on innovative topics, it was possible to prove the importance of ethical and moral barriers when evolving with data-driven digitization. On one hand in case of people analytics it was shown how it can be done right and on the other hand in case of Post Austria it was proven, that ethical and moral standpoints build serious damage to companies when neglected.

6.2 Limitations of the Research

Due to limited time and resource it was hardly possible to cover more aspects of data driven digitization. From my point of view one of the most important ones would have been data/information security, which is covered at a surface level based on basic research.

Further it was also not possible to get more airline data due to concerns about competitors and GDPR issues. The solution approach could have been bigger incl. more maturity and functions if the related data was provided.

Another particular area of research would have been the contrast between agility and data driven innovations, which is also covered on a surface level. In my opinion agility and flat hierarchies often can be the key success factor in many legacy companies and will gain meaning in near future.

3. Implications and Contributions of this work: (Academic contribution & future Research

This research work addresses a holistic approach to data driven digitization. It clearly spares no effort to show, that such approach is everything about doing business in a dynamic environment and understanding the combination of automated, optimized business/technical processes & methods within omni-channel to create sophisticated emotion based customer experiences with a clear ethical standpoint. Through qualitative research this work proves that, there is no such possibility of skipping customer oriented innovation within digitization, because emotion-based innovations backed up by moral and ethical codex is the only way to survive in near future.

From a scientific perspective, this work represents the opinion that related organizational structures as agile hierarchies, technology solutions as Big Data technologies, modern concepts of ecosystem based approaches as lambda/kappa architectures and data/technology governances together can only drive data centric digitization.

One major contribution of this thesis is the practical solution approach to the concept of "Embedded BI", which is a new phenomenon in the market since 2018. Not much research work and scientific solution approach is existing in this very field, which is the reason why the topic was chosen to build a prototype demonstrating how a platform could combine the best of "BI" and "Data Analytics" world for a wide range of consumers within big companies.
The solution approach of so called "Digital Cockpit" from this thesis is all about establishment, operation and visualization of those KPIs by providing a combination of interactive analytics and BI reports for the group executive committees and could be developed further to create new innovations in major industries.

6.1 Final summary

Despite whatever economical and technical evolution might happen in near future, humanity has many major challenges in organizing and running the economy deeply coupled with ethical and sociological questions. For the sake of completeness we have to think about sustainability and take regard of natural resources more than ever not to destroy our planet. Additionally we must be on the right track to optimize & enable innovations face the challenges of 21st century.

Not every innovation as data-driven digitization is maybe fertile, resource saving or even money saving but on the other hand the true power of science and technology is the boundless power of imagination, which should not have any boarders. As there are both merits and demerits in anything, also data-driven digitization has despite all positives many critical and ethical questions yet to answer. In a matter of fact outcomes of surveys prove that there are plenty reasons to be worried such as "job loss through digitization", or even "spying on individuals". At the same time there are also many reasonable facts existing that are proven right through last decades. One particular hypothesis say that every wave of socio economic transformation led through innovation has had enormous positive impacts on mankind.

Even though many short-skilled jobs will vanish from the market in upcoming years, but nevertheless therefore other well-paid jobs will most probably emerge. For instance during 60s & 70s we had e.g. Electronic/Automated Data Processing Managers (EDP/ADP) who became between 80s and now Chief Information Officers (CIOs) taking care of IT departments and their progressions. Since years we are experiencing again a major shifting of this role and related responsibilities. For really long time there was a wall between IT & Business but the exciting journey to digital universe demands more and more all-rounders such as Chief Digital Officers CDOs, Chief Technology Officers CTOs or when it comes to operational roles we now need Data

Officers, Data Scientists, Governance SPOCs or even IT masterminds who have an overall view on Business & IT both fields. Sure specialists on skilled areas will always be indispensable and will probably find skilled jobs easily, but this applies for

More People Are Working Service Jobs
US private-sector employment excluding farm labor
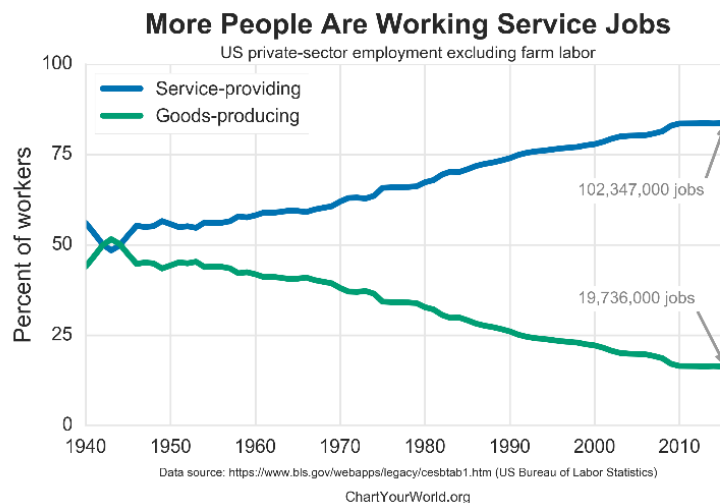
Picture 51 © ChartyourWorld

Data source: https://www.bls.gov/webapps/legacy/cesbtab1.htm (US Bureau of Labor Statistics)
ChartYourWorld.org

developed western markets in for macro-economic context as well. For example, according to bureau of labor statistics website figures from the US show that in February 2017, 145.8 million people in the US had a job - 205 percent more than in 1970, when computers were still in their infancy. The population grew by only 58.8 percent in the same period. Many times scientists, futurists and well—established journalists predicted that automation will kill nearly 90% of the jobs between 50s and now, most of those assumptions and predictions were proven wrong.

As mentioned and justified in this thesis through detailed scientific research on banking and aviation sector, chances are very high that also in case of data-driven digitization something likely will happen. Many other examples, market analysis within the thesis have proven the importance of data-driven digitization in modern decision making & company strategy. In addition to that if matters of ethics, moral and legal standpoints are taken into consideration, it is demonstrated that companies can become more data & tech driven company by creating digital platforms, products to offer added value to the customers, which has physical & emotional components. The survivors of the current shifting phase will be those who transform themselves & learn to run effectively their business as a tech company. Undertaking the right steps to reshape their processes, services but also product portfolio on a constant basis, on a fundament of data-driven digitized business built on ethical management strategy will be the key to success.

6.2 Bibliography of research

[1] **Information Plattform – "**How Much Data Do We Create Every Day?" by Forbes : https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#7712353f60ba

[2] **Scientific Paper** - Analysis of Netflix architecture and business model by Elana Oat (Alto University School of Science) http://www.cse.tkk.fi/fi/opinnot/T-109.4300/2015/luennot-files/oat.pdf

[3] **Scientific Analysis/Journal** - Analysis of Impact of Demonetizationon Digital Transactions in India´- by Dr. Ajay Pratap Yadav1, Dr. Awadhesh Kumar Tiwari (IOSR Journal of Business and Management (IOSR-JBM)

[4] **Website:** The World's Richest and Poorest Countries – Global Finance Magazine (https://www.gfmag.com/global-data/economic-data/worlds-richest-and-poorest-countries)

[5] **Trade Paper -** Aadhaar: Inclusive by Design A Look at India's National Identity Programme and its Role in the JAM Trinity – by GSMA (https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2017/03/gsma-aadhaar-report-270317.pdf)

[6] **Scientific Paper** - "Benchmarking Fast-Data Platforms for the Aadhaar Biometric Database" by Yogesh Simmhan (Department of Computational and Data Sciences, Indian Institute of Science, Bangalore, India

[7] **Journal** Fusion of Disruptive Technologies: Lessons from the Skype Case - European Management Journal Vol. 24 (BHARAT RAO, Polytechnic University, New York BOJAN ANGELOV, Polytechnic University, New York ODED NOV, Polytechnic University, New York) http://faculty.poly.edu/~brao/skype.pdf

[8] **EFMA Report –** https://www.accenture.com/t20171004T035054Z__w__/us-en/_acnmedia/PDF-62/EFMA-Accenture-Awards-2017-WhitePaper-Electronic-09-27.pdf

[9] **Industry White Paper by ORACLE** - The Future of Retail Banking Legacy IT Transformation http://www.oracle.com/us/ciocentral/oracle-legacy-banking-final-395836.pdf

[10] **Industry White Paper by ORACLE** – "Oracle banking platform"
http://www.oracle.com/us/industries/financial-services/oracle-banking-platform-ds-2872623.pdf

[11] **Journal** – "Where Banks are spending their technology dollars" by Center State https://csbcorrespondent.com/blog/where-banks-are-spending-their-technology-dollars

[12] **Scientific Research** – "Big Data Architecture" by Bhashyam Ramesh (Teradata Corporation, Dayton, USA)

[13] **Trage Paper -** 2017 Planning Guide for Data and Analytics by Gartner
https://www.gartner.com/binaries/content/assets/events/keywords/catalyst/catus8/2017_planning_guide_for_data_analytics.pdf

[14] **Scientific Paper –** "Data Lake: A New Ideology in Big Data Era" by Pwint Phyu Khine & Zhao Shun Wang (University of Science and Technology Beijing, China)

[15] **Consulting Ppaper -** How a Modern Data Architecture Will Revolutionize the Financial Services Industry by ClarityInsights
https://cdn2.hubspot.net/hubfs/2106877/White%20Papers/MDA-FS.pdf?t=1527621935324

[16] **Framework Document -** DAMA-DMBOK2 Framework by Patricia Cupoli, Susan Earley, Deborah Henderson (The Data Management Association)

[17] **Scientific Report** – "Alternative approaches to implementing your data lake" by ScienceSoft https://www.scnsoft.com/blog/data-lake-implementation-approaches

[18] **Scientific Bulletin** – "Bulletin of the American Society for Information Science and Technology" by Joan Starr

[19] **Statistics Paper** – "Modelling metadata in central banks" by David Bholat ,European Central Bank) https://www.ecb.europa.eu/pub/pdf/scpsps/ecbsp13.en.pdf

[20] **Book** – "Understanding Metadata" by NISO & ANSI

[21] **Book** – "The Rise of the Meritocracy" by Michael Young

[22] **Book review** - Michael Young's The Rise of the Meritocracy: A Philosophical Critique by Dr Ansgar Allen, University of Sheffield
https://www.sheffield.ac.uk/polopoly_fs/1.155163!/file/philosophicalcritique.pdf

[23] **Management review article** - "Achieving Meritocracy in the Workplace" MITSloan

[24] **Annual report** - by International Air Transport Association of 2018
https://www.iata.org/publications/documents/iata-annual-review-2018.pdf

[25] **Research Report** – "Low-Cost Airlines Bringing the EU closer together" by CEPS (Mehtap Akgüç, Miroslav Beblavý & Felice Simonelli)

[26] **Annual Report** – "The World of Air Transport in 2017" by ICAO
https://www.icao.int/annual-report-2017/Pages/the-world-of-air-transport-in-2017.aspx

[27] **Enterprise Cloud Strategy Paper of 2017** – by Microsoft (Barry Briggs Eduardo Kassner) https://info.microsoft.com/rs/157-GQE-382/images/EN-US-CNTNT-ebook-Enterprise_Cloud_Strategy_2nd_Edition_AzureInfrastructure.pdf

[28] **Annual report** – by Microsoft (2018)

[29] **Scientific Paper -** Datenschutz, Datensicherheit und ethische Aspekte in der Lehrevaluation by Meinald T. Thielsch, Christian Meese und Martin Salaschek Westfälische Wilhelms-Universität Münster, Institut für Psychologie

[30] **Industry Research Paper** – Big Data : Ethische Fragen by Anna Wehofsits (Vodafone institute für Gesellschaft und Kommunikatio

[31] **Report –** „Banks scramble to fix old systems as IT 'cowboys' ride into sunset" by CNBC https://www.cnbc.com/2017/04/11/banks-scramble-to-fix-old-systems-as-it-cowboys-ride-into-sunset.html

[32] **Book** – People Analytics by Ben Weber

[33] **MIT Sloan Report** - People Analytics' Through Super- Charged ID Badges by C. Kane

[34] **GDPR Law website** https://dsgvo-gesetz.de/

[35] **Annual yearly report** – Post Austria (2016)