UNIVERSITY OF TURKU

ABSTRACT

Data warehousing and extract-transform-load-analyze process are an integral part of the Business Intelligence as a Service organization's service offering. However, poor quality of incoming data and client-specific customization coupled with high-growth pose scalability challenges to the case organization's operating model. Moreover, data transformation tasks cannot be fully automated, leading the organization to need new data management resources continuously. When repetitive human involvement requiring data transformation tasks are conducted in high-cost countries, the production costs increase in relation to the new clients taken care of on a continuous basis.

This dissertation focuses on understanding the optimal operating model for executing the repetitive data transformation tasks cost-efficiently without scarifying the organization's superior performance. This qualitative action research follows the IS outsourcing stage model framework and comprises several sub-steps to meet the research aims. The research results demonstrate the root cause for the scalability issues, define the evaluation criteria, and describe the project scope. Consequently, the results evaluate the potential benefits and risks related to current domestic in-house production, outsourcing, and insourcing scenarios. Finally, the research arrives at the operating model recommendation.

The findings underline that the data transformation tasks are an essential part of procurement analytics BIaaS organization's service provision, and therefore they should be kept in-house. Lower cost location allows additional investments in data management capabilities to strengthen the existing core competencies, improve the data transformation output quality, and enhance the overall service level. The results suggest that the case organization should transfer its repetitive data transformation tasks to nearshore or offshore subsidiary to minimize the operating costs and maximize the qualitative benefits.

Turun kauppakorkeakoulu • Turku School of Economics

SELECTING THE BEST SOURCING OPTION FOR A COMPANY OFFERING BUSINESS INTELLIGENCE AS A SERVICE

Action Research in Procurement Analytics and IS Outsourcing

Master's Thesis
in Information System Science

Author:
Saara Gratschev

Supervisors:
Dr. Timo Leino
Dr. Emiel Carol

4.6.2021
Helsinki, Finland

ACKNOWLEDGMENTS

First and foremost, I want to thank my family, Kristiina, Jani, Kaisu, and Arno, for their endless support during my university years. Especially the past two years have been very challenging with worldwide pandemic coupled with the combination of demanding studies and working. Thank you for the encouraging words in the moments of tiredness and frustration. I would not be writing this acknowledgment without you.

Second, I would like to acknowledge my supervisor, prof. Timo Leino, for the gentle guidance and expertise, shared throughout the process. Your support and precious feedback made me trust the process and stay positive in moments of uncertainty.

Third, I wish to acknowledge my employer and the case company *Sievo Oy* for giving me your full confidence to manage this challenging project. I am truly grateful for getting this opportunity to go out of my comfort zone and grow professionally. I also want to express my gratitude and appreciation to my supervisor Perttu Paarma, whose support, feedback, and collaboration have facilitated this thesis project. Furthermore, I want to show appreciation to my manager Nicolas and colleague Tii, who have encouraged me in difficult moments.

Last but not least, I am truly grateful for all the people who had been part of my journey during the six years at the university. Lianne and Manon, thank you for genuinely being there for me during my highs and lows. Fellow IMMITers, I am absolutely proud of us. It was a pleasure to share this adventure with you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI = Artificial Intelligence

BDA = Big Data Analytics

BI = Business Intelligence

BIaaS/SaaS BI = Business Intelligence-as-a-Service

CISO = Chief Information Security Officer

DLP = Data Loss Prevention

DM = Data Management

EEA = European Economic Area

ERP = Enterprise Resource Planning

ETL = Extraction, Transformation, and Load

ETLA = Extraction, Transformation, Load, and Analyze

FTE = Full-Time Equivalent

GDPR = General Data Protection Regulation

HGFs = High-Growth Firms

ICT = Information and Communication Technologies

IS = Information System

KPI = Key Performance Indicator

OLAP = Online Analytical Processing

PO = Purchase Order

PoC = Proof of Concept

R&D = Research and Development

SaaS = Software-as-a-Service

TCT = Transaction Cost Theory

1 INTRODUCTION

1.1 Background

Sievo is a Finnish Business Intelligence as a Service (BIaaS) company grown to be the European market leader in procurement analytics and spend analysis. Continuous high growth results in an increasing amount of repetitive and manual data transformation work as a part of the data warehousing process. When the manual data transformation work takes place in high-cost countries, the cost of labor is high. This thesis aims to investigate other potential operating models by changing the ownership or the location of the manual data transformation work.

New megatrends such as globalization, digitalization, the Internet of Things, and sustainability challenges have increased the expectations towards purchasing. The procurement function is expected to adopt more and more exploratory activities to discover and exploit new knowledge benefiting the whole company. (Gualandris, Legenvre, & Kalchschmidt 2018.) The emergence of new Information and Communication Technologies (ICT) has notably increased the volume of generated data in supply chains (Arunachalam, Kumar, & Kawalek 2018). Moreover, international mergers and acquisitions have impacted procurement, resulting in a growing number of suppliers to be managed. The increasing complexity demands more developed procurement capabilities and matured technologies. (Barrad, Gagnon, & Valverde 2020.) For approaching and solving the problems related to procurement data management (DM), companies must establish an integrated data governance system. This system enables having data from several data sources processed and presented, facilitating superior decision-making. (Handfield, Seongkyoon, & Choi 2019.) To stay competitive, companies must exploit their large data sets wisely. However, many organizations lack the powers of organizing, governing, analyzing, and deploying information assets. (Dallemule & Davenport 2017.)

Developing data transformation capabilities in-house is a time-consuming and expensive process. In addition to challenging development work, organizations often lack the capabilities of conducting the required data transformation tasks such as data cleansing, classification, and enrichment by themselves (Sharda et al. 2015; Limberakis 2012). There has been an increasing tendency to acquire spend analysis from external BIaaS solution providers (Limberakis 2012). In the SaaS business, the more scalable and standardized the software solution is, the less operating costs grow in relation to the

organization's revenue growth, enabling increasing profitability (Sun et al. 2008). Nevertheless, the quality of procurement analytics depends heavily on data transformation activities, which cannot be purely automated. Still, manual human involvement is the key to the high accuracy of the analytics (Maržić et al. 2014). Suppose a BIaaS organization is in high growth and frequently obtaining new clients, but specific data transformation tasks are not scalable. In that case, there is a growing need for an additional data management workforce.

If the data transformation tasks cannot be automated through technical development, facilitating better business model scalability, in that case, the BIaaS organization must research other potential options for improving its profitability in continuous growth. Changing the ownership - *outsourcing* the tasks to an external service provider or *insourcing* - establishing a new entity in a lower-cost country could decrease the total cost of manual data transformation tasks. Nowadays, it is rare not to partly outsource a company's Information Technology (IT) function (Han & Mithas 2014). The most typical reasons for outsourcing are cost-savings, freeing resources focusing on core competencies, and enabling operational flexibility. Outsourcing or insourcing the heavy repetitive data transformation tasks could increase the BIaaS business model's scalability. However, despite the potential gains, there are always trade-offs in outsourcing and insourcing solutions. Changing the ownership or location of the data transformation work can potentially occur a wide range of potential risks, such as increasing information security vulnerability, loss of management control, and employee morale issues. (Yang et al. 2007.) Outsourcing and insourcing expose the BIaaS company to a large scale of potential emerging benefits and risks. There is a need for an in-depth understanding of how the company can efficiently execute these repetitive data transformation tasks without sacrificing its strong growth enabling success factors.

The paper follows an action research approach using IS Outsourcing Stage Model (Dibbern et al. 2004) as a theoretical framework. Action research projects are situation-specific; their objective is not to generate universal knowledge (Coughlan & Coghlan 2002). However, this research provides an example of how the BIaaS organization is guided through the three main steps of IS outsourcing research phase: "*why*," "*what*," and "*which*." A set of different qualitative methods are exploited for defining why to outsource or insource, what to outsource or insource, and which outsourcing or insourcing option to choose.

1.2 Introduction to the case company

Sievo is a Finnish procurement analytics SaaS solution provider founded in 2003. Their primary offering is an online analytical processing procurement information hub consisting of several solution modules. Those modules are spend analysis, savings lifecycle, materials forecasting, contract management, and procurement benchmarking. Their solutions use data warehousing architecture and offer especially high-profile extract-transform-load-analyze (ETLA) services. The customer journey starts with the project-based implementation phase, where clients' historical purchasing data is extracted, migrated, classified, enriched, configured, and visualized as a bulk. Once the software is implemented, the existing solution is continuously maintained and developed. The clients transfer their latest procurement data once a month or a week to Sievo's database, and it is similarly data warehoused. Despite the Finnish background, the company has mainly served large multinational corporations from different sectors. For example, Carlsberg, ISS, SNCF, and Schindler can be found in their client portfolio. Sievo's strategy is to be a top-end solution with above-average pricing. It also means that the clients have high expectations towards data treatment, data quality, and customer service.

Sievo is divided into eight functions: People & IT, Finance, Existing Customers, Professional Services, Sales, Marketing, Engineering, and Product Management. Figure 1 demonstrates how *data transformation activities* occur in two separate data management teams in Existing Customers and Professional Services functions. For clarity, later in this paper, those two functions are called "continuous service function" and "implementation function." DM Specialists take care of repetitive data transformation tasks and more complex development work in both functions. When it comes to repetitive work, continuous service DM Specialists are in charge of processing the client's new procurement data on a monthly or weekly basis. Additionally, they make corrections to existing historical data based on the clients' feedback. Implementation DM Specialists instead are responsible for the implementation of the solution for the new clients. The work is executed on a customer-by-customer project basis. The function consists of several project teams led by project managers and assisted by solution consultants and data engineers. Different project teams then utilize the available resources of the separate DM team for processing all the clients' historical data in bulk. Moreover, DM Specialists of the

implementation function assist the sales function with proofs-of-concept (PoCs). Sometimes at terse notice, Sievo must demonstrate its data warehousing capabilities to its prospects.
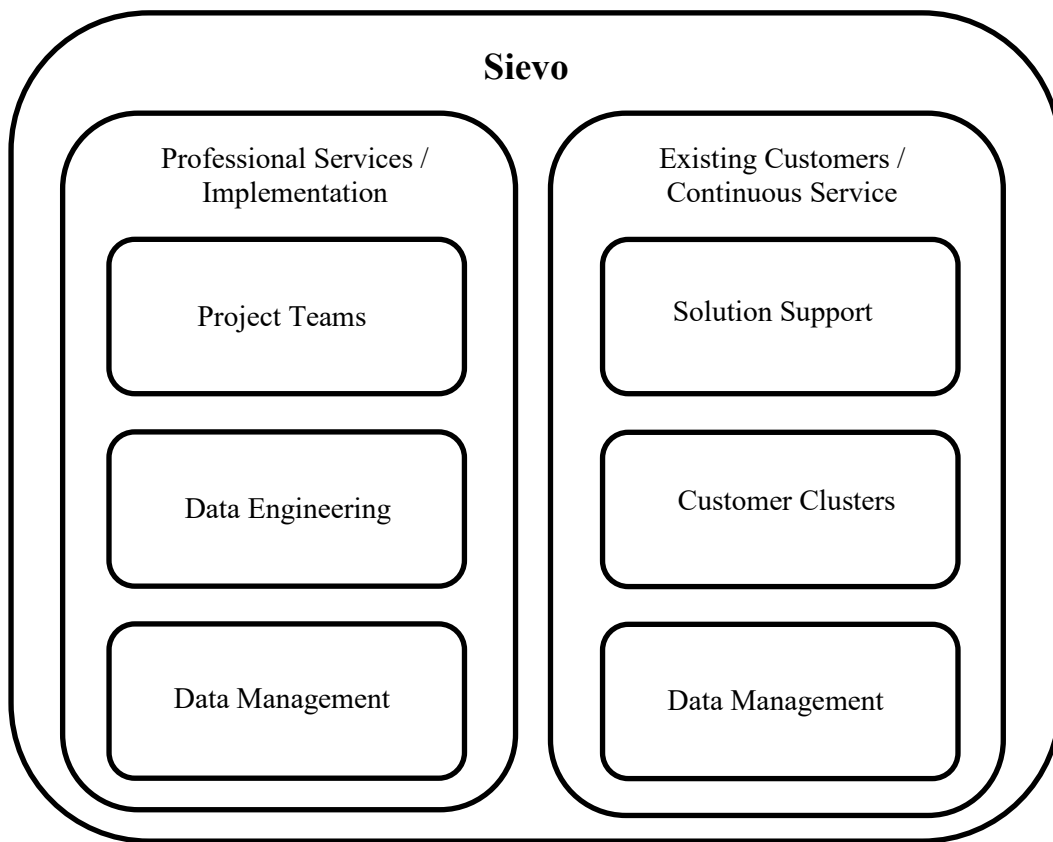
Figure 1 Data management across Sievo

1.3 Research question

The aim of this dissertation is to find the optimal operating model for executing repetitive procurement analytics data transformation tasks. In order to meet the research objective, the paper aims to cover the following research question:

1. What is the optimal way of organizing manual, repetitive data transformation tasks in a Business Intelligence as a Service organization?

2 THEORETICAL BACKGROUND

2.1 Business Intelligence as a Service for procurement analytics

First, the concepts of Business Intelligence, Procurement Analytics, and Software as a Service are covered. Additionally, the chapter explains how high-growth impacts the SaaS business model and what are the scalability challenges.

2.1.1 Business Intelligence

The concept of Business Intelligence (BI) has a dual meaning. First, it refers to data analysis that exploits large data sets to understand organizational activities and opportunities, enabling more accurate strategic decision-making. Secondly, BI denotes a set of tools supporting the data analysis, e.g., querying, scenario modeling, and dashboarding. (DAMA, 2017.) Figure 2 demonstrates BI facilitating data warehouse architecture. It consists of different components: data sources, ETL process, data warehouse, data marts, middleware tools, and front-end applications. The data warehouse is fed by data extracted from several internal and external sources. Internal sources are typically ERP and legacy systems. In contrast, external sources are external service providers providing additional data points, e.g., Moody's for credit ratings and Dun & Bradstreet for organizational information.

After the data extraction follows the other two main steps of the ETL process – transformation and loading. (Sharda et al., 2015.) Data transformation refers to the means of increasing the data quality by altering the structure and format of data. Different forms of data transformation are data translation, data classification, data consolidation, and data encryption. (Nadikattu, 2019.) After the transformation actions, the integrated and cleansed data is loaded to the warehouse optimized for data analytics (Sharda et al., 2015). The data warehouse is a central repository of integrated data from one or more data sources. It usually holds replication and historical copies of data facilitating time-based analysis. The nature of data in the warehouse is rather "aggregated" or "summarized" than atomic. (van Gils, 2020.)

Metadata defines the data available in the warehouse based on certain specifications such as data origin, functionality, value, and characteristics (Nadikattu, 2019). Data marts are used in situations where different stakeholders have different requirements for the same BI data. Data marts hold a subset of the data in the warehouse, which is tailored

to correspond to the needs of a specific stakeholder group, such as the marketing or finance department. (van Gils, 2020.) Middleware tools, such as SQL or more managed querying environment Business Objects, enable access to the data warehouse for experienced users. Business users can interact with the data through front-end applications, allowing multiple actions such as routine reporting, data mining, online analytical processing (OLAP), and custom-built services. (Sharda et al., 2015.)
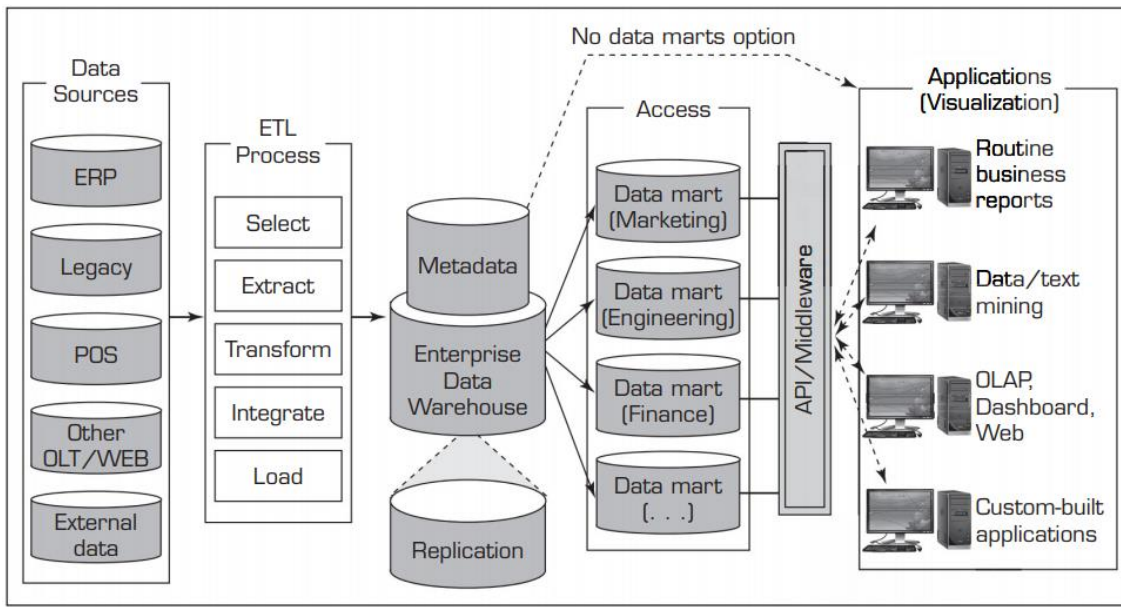
Figure 2 A data warehouse framework and views (Sharda et al. 2015)

2.1.2 Procurement analytics

Procurement analytics can be considered as offensive DM activities whose main objective is to improve an organization's competitive position and profitability (Dallemule & Davenport 2017). Typical procurement analytics problems and decisions are connected to spending and budgets, cost reduction and modeling, supplier management and evaluation, category market intelligence, procurement-driven innovation, market strategies, supply chain risk, and stakeholder value improvement (Monczka et al. 2016). The most requested procurement analytics area is spend analysis, followed by price benchmarking, supplier performance management, and risk alerts (Handfield et al. 2019). Indeed, spend analysis and cost visibility are an integral part of procurement analytics, enabling identifying new business openings and cost savings. (Barrad, Gagnon, & Valverd 2020.) Spend analytics platforms combine the purchasing data with user data, business unit origin, and supplier information. Typically, these technologies facilitate analytic insights by integrating database querying capability and visualization graphics. (Handfield et al. 2019.)

According to Maržić, Krneta, & Pavlic (2014), spend analysis signifies organizing spend data by exploiting supplier and product hierarchies to discover:

1) Spend categories
2) Strategic sourcing opportunities by
   a. collecting procurement requirements
   b. supplier consolidation
3) Cost reduction possibilities due to increased information and bargaining power

Typically, the transactional system contains data on a specific purchase, but it does not provide transactions in an aggregated format. By finding spend categories for goods and services, an organization can analyze spend distribution on multiple levels. For example, the IT department can solely exploit the IT equipment expense category and its subcategory transaction data. When an organization has a good overview of its spending, it can gather a set of requirements that make the procurement more efficient and cost-effective, guiding the sourcing to a more strategic direction. Small orders of one single supplier can be aggregated into larger ones, or complementary products and services can be collected in the same order. (Maržić et al. 2014.)

Furthermore, strategic sourcing opportunities are discovered through supplier rationalization, with other terms supplier consolidation, inferring to identifying the vendors that are part of the same corporation. (Maržić et al. 2014.) Often different business units within large global organizations unwittingly acquire the same goods or services from various service providers or even from the same global supplier with different prices. Supplier consolidation aids the organization in recognizing comparable goods or services that could be standardized and aggregated. (Monczka et al. 2016.) With these analytical insights, it is possible to obtain a bargaining power over vendors enabling cost reductions through higher discounts (Maržić et al. 2014).

Procurement analytics cannot be solely generalized as equal to the ETL process and analytics (ETLA). Still, the strategic advantages emerge from the capability to exploit spend analysis as a predictive means for improving spend compliance and decreasing supplier risks (Limberakis 2012). However, the quality and accuracy of spend analysis depend heavily on the quality of the ETL process, more specifically, of the data *transformation* step. High-quality data transformation forms a firm ground for further analysis; thus, it is the most critical part of the spend analysis process. Figure 3 showcases the data transformation process in the context of spend analysis. The data transformation step can

be considered as a process consisting of five sub-steps: normalization, enrichment, classification, supplier consolidation, and learning. (Maržić et al. 2014.)
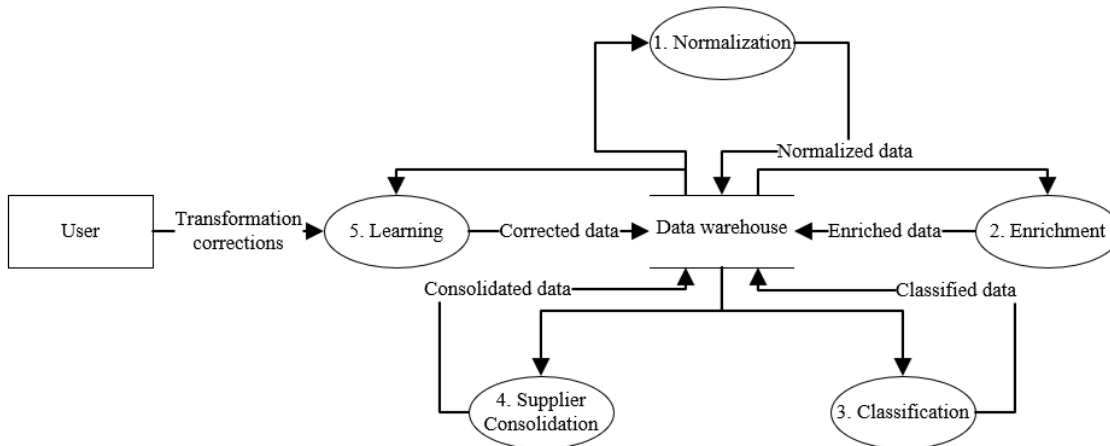
Figure 3 Data transformation process diagram (adapted from Maržić et al. 2014)

Normalization refers to the tasks of structuring data corresponding to a specific ordering. Additionally, it might require converting different currencies and units of measure to a comparable format. After that follows *data enrichment* which means increasing the quality of the data and potential insights by filling missing data and conjoining data from external data sources such as stock market or supplier financial data. *Supplier consolidation* implies creating supplier hierarchy by connecting subsidiary suppliers with the mother company allowing calculating the total spend of a global supplier rather than its one branch office. (Maržić et al. 2014.) Data *classification* denotes assigning a label, which defines a specific class, to a spend object. In the context of spend analysis, the classification is usually implemented as supervised machine learning, meaning that the classes are based on a predefined class structure – taxonomy. There are multiple classification algorithm categories such as statistical algorithms, decision tree-based algorithms, and rule-based algorithms. (Pandit & Marmanis 2008.) Yet, classification and consolidation algorithms are not absolutely flawless, requiring human involvement in correcting data transformation results and enabling continuous *learning* towards better accuracy (Maržić et al. 2014).

2.1.3   Software as a Service

Software-as-a-Service (SaaS) partakes in a more expansive "servicification" phenomenon which refers to turning products into tools that are owned, maintained, and updated by third-party service providers (Sturgeon 2021). SaaS can be described as a provisioned

software over the internet and utilized commonly with a web browser (Mäkilä et al., 2010). Typically, a SaaS application is developed with highly standardized functionalities so that the vendor could exploit economies of scale and serve as many customers as possible. Its main advantages are the lower total cost of ownership and increased mobility. (Sun et al. 2008.) Additionally, the on-demand model provides the clients greater flexibility, increased security, and shorter time-to-market than internal development, resulting in the industry's notable growth during the past decade worldwide (Deloitte 2021; Statista 2021).

The growing popularity of servicification coupled with data-driven business management has led BI market offerings to shift towards BIaaS solutions (Liyang et al. 2011). BIaaS products enable the end-users to access the data through a web interface. Moreover, it liberates customer organizations from data warehousing implementation and continuous maintenance responsibilities. Service providers are in charge of guaranteeing the data warehouse platform's performance, availability, and scalability. When corporations grow, and their business processes become broader and more complex, BIaaS serves as an efficient measure orchestrating different business processes and systems. (Sano 2014.)

As seen in the previous chapter, correctly performed spend analysis provides the organization increased visibility, improved decision-making, and cost-saving opportunities. Still, many companies are lacking DM skills for executing required spend analysis in-house. A critical success factor is the ability to access, organize, and analyze spend data accurately. (Limberakis 2012.) Developing data transformation capabilities in-house is a time-consuming and expensive process, but at the same time, online analytical processing is dependant on data transformation quality (Sharda et al. 2015). There has been an increasing tendency to acquire spend analysis from external service providers due to a lack of skills in conducting data analysis tasks internally, such as data cleansing, classification, and enrichment (Limberakis 2012). These factors make the acquisition of procurement analytics BlaaS solution from external service providers an attractive option.

2.1.4 Challenges of high-growth for BIaaS providers

High-growth firms (HGFs) have drawn the attention of academics and policymakers already for several years due to the extraordinary nature of companies' performance metrics (Coad et al. 2014). OECD (2007) defines HGS as the following: "All enterprises with average annualized growth greater than 20% per annum, over three years, and with ten or more employees at the beginning of the observation period. Growth is thus measured by

the number of employees and by turnover". HGFs create a disproportionately large number of new jobs (Coad et al. 2014). Knowledge-intensive IT service providers are overrepresented amid high-growth enterprises (Asikainen & Mangiarotti 2017).

According to Yen Ng & Hamilton (2016) the main drivers for high growth in technology-based companies are customer-focused flexibility, commitment to research and development (R&D), and employee engagement. Firstly, HGFs are increasingly expected to be flexible, answer their customers' expectations efficiently, and offer custom solutions to sustain their growth. Secondly, to be capable of fulfilling those expectations, HGFs must heavily invest in R&D. Innovation capability enables satisfying customer requirements, but also it is an asset, which competitors do not easily emulate. Lastly, in technology-intensive HGFs human resource practices are one of the critical drivers of growth. By nurturing strong employee engagement and an open company culture, an organization is proven to be more responsive and innovative. (Yen Ng & Hamilton 2016.) Additionally, Lopez-Garcia & Puente (2012) found a positive correlation between wage premium paid compared to other firms operating in the same sector and the probability of fast growth. Moreover, the higher number of permanent workers in contrast to temporary workers has positively impacted the organization's growth (Lopez-Garcia & Puente 2012).

Even though companies are constantly pursuing high growth, it also presents challenges in the BIaaS context. Solely SaaS business model in high-growth poses a scalability threat. When strong growth is facilitated by customer-focused flexibility, meaning fulfilling customer-specific configuration and customization requirements, the possible scalability immediately decreases (Sun et al. 2008). Furthermore, each new customer generates additional implementation and operating costs of the technological infrastructure required to deliver the SaaS application. At the organizational level, economies of scale are complicated to apply to continuous account management and customer service when customer orientation is one of the key drivers of high growth. (Novelli 2011.)

When it comes to BI and procurement analytics, classification algorithms are not enough matured to provide completely accurate results; therefore, human involvement is required to correct data transformation results and make reclassification decisions (Maržić et al. 2014). The process's high importance demonstrates that ETL usually comprises 70% of the total time of data-intensive IT projects (Sharda et al. 2015). Human is capable of detecting errors, specifying cleaning rules, and producing feedback for machine learning algorithms. However, human input is expensive, and organizations have

limited resources for cleaning work. (Abedjan et al. 2016.) Manual editing is not scalable with respect to the size of the data. The speed of manual processing is multiple orders of magnitudes smaller than automated processing. (Pandit & Marmanis 2008.) The combination of continuous intensive growth and the aim of fulfilling high customer expectations signifies that human involvement in ETL is not scalable, and the enterprise is opposed to constantly acquiring additional workforce. Effectively, it instead leads to the situation where the costs of the operating model increase with the organization's revenue.

2.2   Information Technology outsourcing and insourcing

This chapter introduces the concepts of outsourcing and insourcing. After that, the IS outsourcing framework and its sup-steps are discussed in detail.

2.2.1   IS outsourcing framework

Information Technology (IT) outsourcing refers to transferring IT hardware and software, leases, staff, and managerial service delivery responsibilities for external service providers (Hirschheim & Lacity 2000). Outsourcing means allocating or reallocating business activities from an internal operator to an external, whereas insourcing is about changing the geographical location of internal resources (Schniederjans, Schniederjans, & Schniederjans 2005). Insourcing comprises the practice of reducing external contract volume and creating or altering existing governance structures (Foerstl, Kirchoff, & Bals 2016). IT outsourcing, in general, is a widely researched topic, but there is lesser information about the relative strategic value of IT outsourcing and IT insourcing. However, outsourcing and insourcing should be considered equal alternatives for reorganizing IT activities from the organizational perspective. (Qu, Oh, & Pinsonneault 2010.) For comparing the IT outsourcing and IT insourcing scenarios, this research exploits the IT outsourcing stage model by Dibbern et al. (2004) as a theoretical framework.

Figure 4 showcases the information system (IS) outsourcing framework divided into two main stages: decision process and implementation (Dibbern et al. 2004). Three subactivities in the decision process stage are "*why,*" "*what,*" and "*which*." The outsourcing process starts with understanding the root causes and *why* an organization considers outsourcing its data transformation activities. This step also involves analyzing the potential advantages and disadvantages related to outsourcing. Those findings can be

exploited for defining *what* to outsource. The organization must understand the feasible degree of ownership and outsourcing. Once the reasoning behind the outsourcing decision and potential project scope is determined, the next step is to examine *which* choice to make. It means conducting a set of procedures for arriving at an outsourcing decision. There is no one specific guideline to follow, but typically organizations establish or choose existing criteria for selecting the best outsourcing scenario for their purpose. (Dibbern et al. 2004.) The research scope is limited solely to the decision process stage, and therefore the activities related to the implementation stage are not given attention in this paper. The decision process stage is closer discussed in the subsequent chapters.
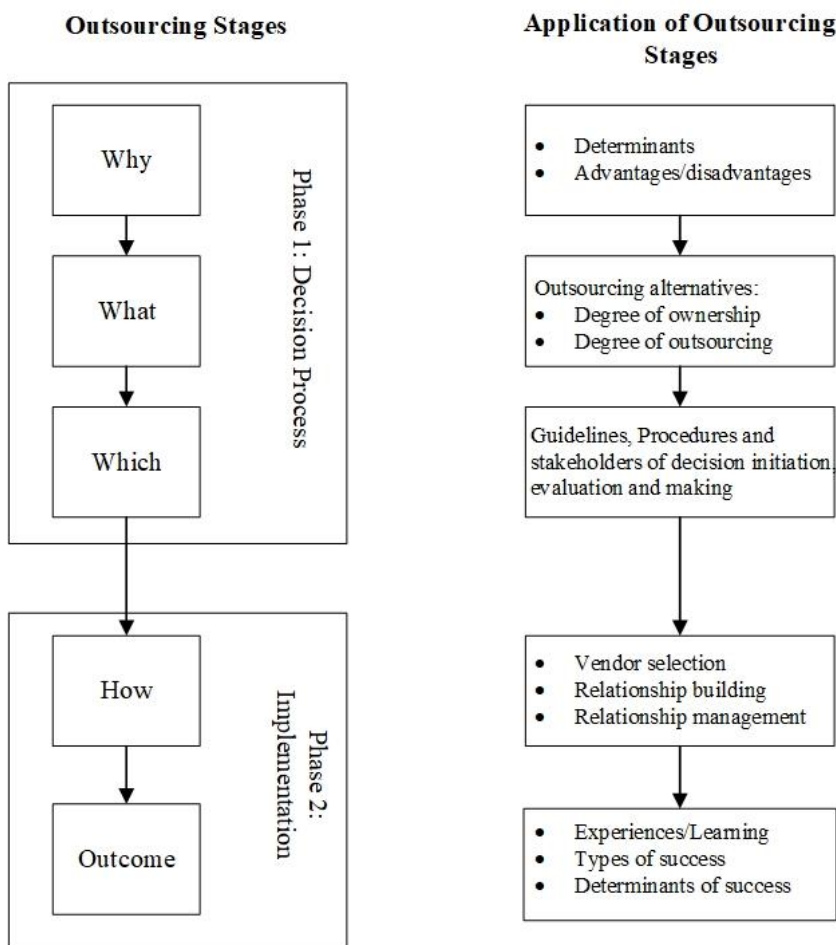
Figure 4 Stage model of IS outsourcing (adapted from Dibbern et al. 2004)

2.2.2   Why: understanding the benefits and risks

IS outsourcing process begins with understanding the rationale why a company considers outsourcing its IS function. It is vital to comprehend the context and the motivation for outsourcing projects. Moreover, the organization should evaluate the advantages and risks related to outsourcing. (Dibbern et al. 2004.) Previously, it was demonstrated that strong

growth poses scalability challenges to the SaaS business model. Furthermore, in procurement analytics, ETL process and data transformation are demanding tasks, requiring a continuously growing amount of human involvement in high-growth organizations. In order to control its cost of labor and increase operational scalability and flexibility, the organization must search for alternative options for executing human involvement requiring data transformation tasks. High-growth companies are prone to face difficulties determining which kind of organizational transitions or changes are required to take due to increased managerial complexity (Nicholls-Nixon 2005). However, outsourcing or insourcing could serve as a solution for emerging challenges in the BIaaS model (Oshri, Kotlarsky, & Willcocks 2009). It is essential to scrutinize the underlying factors. Outsourcing should be considered only when it makes good business sense. (Hodosi & Lazar 2013.)

Outsourcing, in general, may provide several advantages and risks for an organization. Table 1 showcases how potential advantages and risks are translated into criteria for evaluating different outsourcing scenarios. The benefits and risks can be divided into two categories: quantitative and qualitative. From the quantitative perspective, typically, outsourcing may facilitate notable cost savings due to lower production costs. Other savings may emerge through improved efficiency, established economies of scale, and cutting or stabilizing overhead expenses. (Dhar & Balakrishnan 2006; Oshri et al. 2009.)

From the qualitative perspective, outsourcing allows companies to focus on their core business and main strategical objectives. (Dhar & Balakrishnan 2006; Oshri et al. 2009.) It might provide access to a third party's unique field-specific expertise - insights that the client organization can exploit as a competitive advantage (Oshri et al. 2009; Mohr et al. 2011). When the organization experiences rapidly changing demand conditions, such as high growth, outsourcing offers flexibility to scale the production capability at a lower cost. (Oshri et al. 2009.) Additionally, it can reduce time to market (Dhar & Balakrishnan 2006). When outsourcing allows performing the task in a superior manner, it positively contributes to the end-customer satisfaction (Dhar & Balakrishnan 2006; Bairi & Murali Manohar 2011).

Despite the several advantages, outsourcing involves several risks and disadvantages. Effectively, in a longitudinal study conducted by Delen et al. (2016), only 60% of IT-outsourcing projects were found to be successful. Giving control of specific tasks to external service providers might result in loss of critical skills or overdependence on an outside organization. (Mohr et al. 2011.) Organization is prone to opportunism if the

outsourcing involves excessive, uncontrollable switching costs, complex interdependencies between business processes, or lack of supplier performance measurement (Mclvor 2008). The external service provider may fail to deliver the service as agreed (Dhar & Balakrishnan 2006). The overall quality of the procurement analytics solution depends heavily on the quality of data transformation; therefore, there is a high risk of potentially losing control over the timing and quality of the data transformation outputs. (Oshri et al. 2009.) Outsourcing can damage the client organization's customer perception and satisfaction if the external service provider disregards the end-customers (Bairi & Murali Manohar 2011).

Additionally, it is essential to understand the importance of outsourced components to the overall product offering (Hätönen, 2010). If an organization decides to outsource the critical differentiator in customers' eyes, there is a risk of dilution of competitive advantage over time (Mclvor 2008; Mohr et al. 2011). Indeed, outsourcing increases the risk of losing critical skills (Chou & Chou 2009). If the company can perform the task in a unique superior manner compared to its competitors and external providers, outsourcing can have detrimental negative consequences to the organization's competitive position (Mclvor 2008).

IT-related tasks and processes are often under continuous, even unanticipated, technological development, which might pose challenges in outsourcing. It is possible that the vendor cannot support future technological development or the vendor and client do not have matching interests. Technological development can dramatically change the outsourcing needs in the course of time. Even though outsourcing is expected to bring cost savings and improved efficiency, it also requires increased managerial capability. The client organization must invest in contract and relationship management with the external vendor. Unexperienced organizations tend to underestimate the necessary effort to manage changing IT capability, IT governance, and outsourcing projects. (Willcocks, Fitzgerald, & Feeny 1995.)

Finally, it is also vital to understand the consequences of outsourcing to company culture and information security. Outsourcing may hurt organizational culture by demotivating employees, which leads to a loss of employee morale and productivity (Willcocks et al. 1995; Chou & Chou 2009). Moreover, outsourcing data transformation tasks to a third party can raise security and confidentiality-related questions (Dhar & Balakrishnan 2006; Oshri et al. 2009).

Table 1 Advantages and risks translated into evaluation criteria.

Advantage/Risk	Criteria	Reference
+/- Cost savings, new emerging costs	Cost	(Willcocks et al. 1995). (Dhar & Balakrishnan 2006), (Oshri et al. 2009)
+ Efficiency	Efficiency	(Dhar & Balakrishnan 2006),
+/- Higher/poorer performance quality	Quality	(Dhar & Balakrishna 2006), (Oshri et al. 2009)
+Improved operational flexibility, reduced time to market	Required scalability	(Dhar & Balakrishnan 2006), (Oshri et al. 2009)
+Allow focusing on core activities	Need for resource optimization	(Dhar & Balakrishnan 2006), (Oshri et al. 2009)
+/-Accessing/Losing highly skilled resources	Required expertise (technology, industry, customer-specific)	(Chou & Chou 2009), (Oshri et al. 2009), (Mohr et al. 2011)
+/-Strengthening/dilution of competitive advantage	Contribution to competitive advantage	(Oshri et al. 2009), (Mclvor 2008), (Hätönen, 2010), (Mohr et al. 2011)
+/-Higher/lower customer satisfaction	Customer perception	(Dhar & Balakrishnan, 2006), (Bairi & Murali Manohar 2011)
+/- Enabling/losing a unique way of performing	Relative capability position	(Mclvor 2008), (Mohr et al. 2011)
-Misalignment with technological development	Technological development capability	(Willcocks et al. 1995).
-Loss of employee morale and productivity	Impact on company culture	(Willcocks et al. 1995). (Chou & Chou 2009).
-Threat of overdependency, lock-in	Potential for opportunism	(Mclvor 2008), (Mohr et al. 2011)
-Increased vulnerability	Security	(Dhar & Balakrishnan 2006), (Oshri et al. 2009)
-Increased delivery instability	Sensibility to instability	(Dhar & Balakrishnan 2006), (Oshri et al. 2009)
-Increasing management work	Required management capability	(Willcocks et al. 1995).

One of the critical outsourcing success factors is the organization's ability to quantify and qualify its outsourcing needs at the beginning of the project (Hodosi & Lazar 2013). Organizations are often over-emphasizing short-term cost savings in their decision-making. According to Willcocks et al. (1995), the most compatible circumstances for outsourcing are when the impact on business position, level of IT integration, in-house versus market expertise, and link to business strategy are low. Furthermore, high technology maturity and low business uncertainty are in favor of outsourcing. However, it is crucial to understand how the decision impacts the organization's unique long-term capabilities. (Mclvor 2008.) Therefore, it is essential to evaluate other additional factors

than solely the ones defining the optimal outsourcing circumstances (Willcocks et al. 1995). Correspondingly, the organization cannot make outsourcings decision purely based on the short-term cost reduction advantages, but it should rather evaluate different tasks against comprehensive criteria that consider potential emerging benefits and risks.

2.2.3   What: defining the project scope and the scenarios

2.2.3.1   Project scope: the degree of outsourcing

When making the sourcing decisions, an organization must define the project scope, which means determining the degree of outsourcing. The degree of outsourcing can variate from the entire IT department to a specific selected function. (Dibbern et al. 2004.) It is essential to carefully analyze the scale and scope of the activity to be potentially outsourced since the characteristics directly impact the decisions to be taken later in the process (Hätönen 2010).

The degree of IT outsourcing is about deciding between selective and total outsourcing. Selective IT outsourcing is about researching and choosing the most appropriate activities to be outsourced. Total outsourcing instead implies outsourcing the majority of the organization's IT activities. It is also possible that the organization starts with very selective outsourcing, and in the course of time, the organization increases the degree of outsourcing. (Dahlberg & Lahdelma 2007.) Multiple inducements are impacting the degree of outsourcing. Typically, financial factors, the level of production and transaction costs, are prevalent aspects affecting the degree of IT outsourcing. The greater perceived outsourcing cost advantages are, the more significant proportion of the organization's IT assets are outsourced. Furthermore, usually smaller organizations tend to have a higher degree of outsourcing because it is challenging to create economies of scale in IT operations advocating internal IT assets. (Ang & Straub 2006.)

In addition to the degree of outsourcing, the organization must define the degree of ownership of the sourced asset (Dibbern et al. 2004). Figure 5 demonstrates that potential outsourcing and insourcing scenarios can be modeled to three times three matrix: the x-axis represents location dimension, and the y-axis refers to the contractual/ownership dimension. Different project scopes have different optimal combinations of ownership and location. Typically, value-creating activities are most effectively executed inhouse and onshore. In contrast, some other tasks, such as component assembly, are optimally produced in an offshore location by an external partner. (Foerstl et al. 2016.)

Contractual/Ownership dimension

|	Domestic	Nearshore	Offshore
Buy	1. Domestic Sourcing	2. Nearshore Outsourcing	3. Offshore Outsourcing
Hybrid	6. Domestic Partnership	7. Nearshore Partnership	4. Offshore Partnership
Make	9. Domestic In-house Production	8. Nearshore In-house Production	5. Offshore In-house Production

Location Dimension
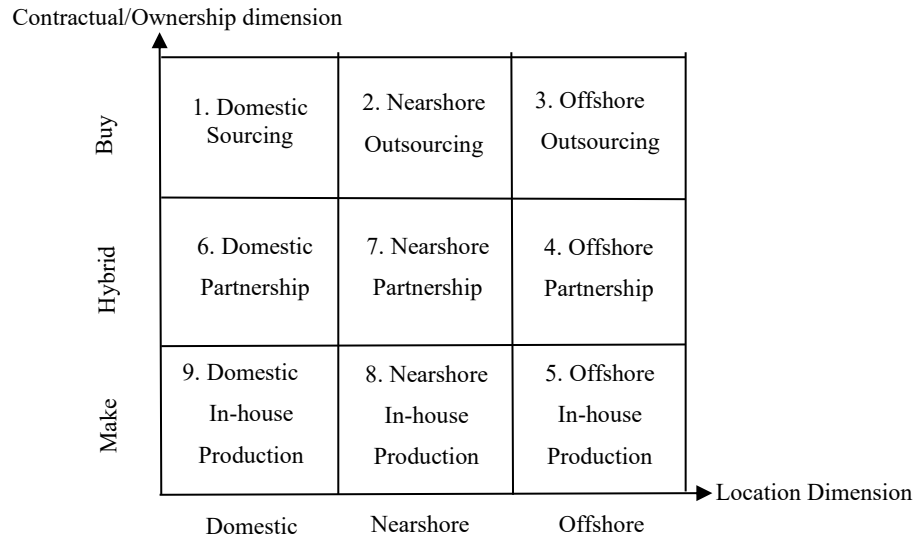
Figure 5 Outsourcing and insourcing alternatives (adapted from Foerstl, Kirchoff, & Bals 2016)

The degree of ownership is divided into three primary levels: buy, hybrid, and make. '*Buy'* indicates regular outsourcing - handing over the ownership of the chosen components or services to a third-party provider. '*Hybrid*' solution means shared ownership of a specific task in the form of a joint venture, strategic partnership, or long-term contract. '*Make'* instead implies full ownership of the IT functionality, which can be arranged as a subsidiary or acquired business unit. (Foerstl et al. 2016). When comparing the advantages and disadvantages of outsourcing and insourcing, typically, labor costs are higher in insourcing than outsourcing options. However, insourcing provides some additional benefits. It reassures the control over production output. Additionally, the organization's workforce is more motivated and loyal than external service providers (Schniederjans et al. 2005.)

Besides the degree of ownership dimension, Foerstl et al. (2016) also underline the relevance of the location when evaluating different potential outsourcing and insourcing scenarios. Location dimension refers to the distance between the organization's home nation and the location of the outsourced or insourced operations. The location options variate amid domestic, nearshore, and offshore production. (Gerbl, McIvor & Humphreys 2016.) When choosing the optimal location, the organization should at least consider the total costs, availability of skills, operating environment, quality of infrastructure, risk profile, and market potential (Oshri et al. 2009). Globally intensifying competition has led the companies searching more cost-effective resources from the less expensive offshore locations such as India, China, and the Philippines (Hirschheim & Dibbern 2006). Even

though the production cost advantages are usually inferior in nearshoring, it still facilitates other types of benefits such as lower travel coss, fewer time-zone differences, and lesser cultural differences than offshore locations.

2.2.4 Which: arriving at the decision

The first prerequisite for sourcing decision analysis is having a description of disaggregated elements of the IS in question. As the second prerequisite organization should define the criteria for evaluating the performance of the identified IS activities. The evaluation criteria should reflect the organization's preferences regarding different metrics, e.g., cost, strategic contribution, and profit generation. Once these two prerequisites are fulfilled, the organization adopts a set of procedures for arriving at the final sourcing selection. There is no grounded theory or standard agreement in academia concerning evaluation criteria and final sourcing decisions. Still, usually, the process consists of initiation, analysis or evaluation, and actual decision-making. Transaction cost theory (TCT) can be used for guiding the decision process. (Dibbern et al. 2004.)

TCT is widely exploited in academic research for explaining and estimating IT outsourcing decisions and outcomes (Alaghehband, Rivard, Wu, & Goyette 2011). TCT serves the client organization's objectives to consider economic theory and organizational reality – not making sourcing decisions based solely on vendors' capability to reach economies of scale through specialization. According to TCT the main drivers for sourcing decisions are minimizing total cost and maximizing total value. Total cost is relatively simple to define, and it can be divided into production and transaction costs. (Lacity & Willcocks 2009.) *Production costs* cover all the expenses related to producing the selected IT services, whereas *transaction costs* are divided into three subcategories:

1) The contract establishment expenses finding vendor, negotiations, legal fees, and other labor incurred.

2) The cost of monitoring and coordinating the activities produced by the vendor

3) Switching cost in case of changing the vendor due to under-performance or failure

Transaction cost theory assumes that the decision-makers are bounded rationality, meaning that they make choices based on their best knowledge, but it is still not the best market information. It sometimes leads to incorrect decisions increasing the transaction costs. Moreover, the TCT expects that different parties of the transactions might show opportunistic behavior requiring additional investments to monitor the counterparty's performance. (Thouin, Hoffman, & Ford 2009.)

Measuring the quantitative expenses is relatively more straightforward task than estimating the potential added value related to outsourcing. There is no one accepted model for measuring the total value of the sourcing project. (Ngwenyama & Bryson 1999.) However, Ngwenyama & Bryson (1999) suggest weighting different value attributes, such as accuracy, timeliness, and reliability, according to the organization's priorities for measuring values for each option considered. As stated by transaction cost economies, organizations tend to choose insourcing over outsourcing on the assumption of too high risk of market failure. Significant firm-specificity in required human assets and a high level of uncertainty are the factors increasing the market failure risk. (Dibbern & Heinzl 2006.)

3 RESEACH METHODOLOGY

3.1 Introduction

Business research is typically related to practical problems connected to an organization's marketing, financing, development, and growth (Eriksson & Kovalainen 2011). Research questions emerge from these real-life issues, and from that perspective, this paper was not an exception. The main objective of this research was to understand the most optimal way of organizing repetitive data transformation tasks in the BIaaS solution provider context. To reach that objective, the current state of the ETL process, the root causes, and potential solutions were required to be analyzed in-depth. The research focused not only on solving the organization's practical business case but also on contributing to existing IT outsourcing literature by providing new insights from the BIaaS perspective.

First, this methodology section provides details on the chosen research strategy. Then different steps of the action research life cycle are covered. Finally, the limitations of the selected research strategy are explained.

3.2 Research strategy

3.2.1 Chosen strategy

The chosen research strategy was action research because the researcher was not only supposed to study the organizational phenomenon but also conduct phase 1 of the IT outsourcing process (Figure 4) by Dibbern et al. (2004) in practice. In action research, the researcher aims to solve a current practical problem while also expanding scientific knowledge. Action research differs from other research methods because the researcher is expected to create organizational change in collaboration with the research subjects. (Baskerville & Myers 2004.) According to Baskerville (1999), there is no one rigid action research method, but it is rather a class of research approaches, which have four common characteristics:

1) focus on action and change
2) problem-oriented approach
3) "organic" research process including systematic and occasionally iterative stages
4) collaborative operating style

As stated before, there is no one fixed model for describing the action research life-cycle (Baskerville 1999). Figure 6 by Coughlan & Coghlan (2002) presents the action research cycle that comprises three types of steps:

1) A pre-step for understanding the research problem and the context
2) six main steps representing data and feedback gathering, data analysis, action planning, implementation, and the evaluation
3) continuous monitoring as a meta step
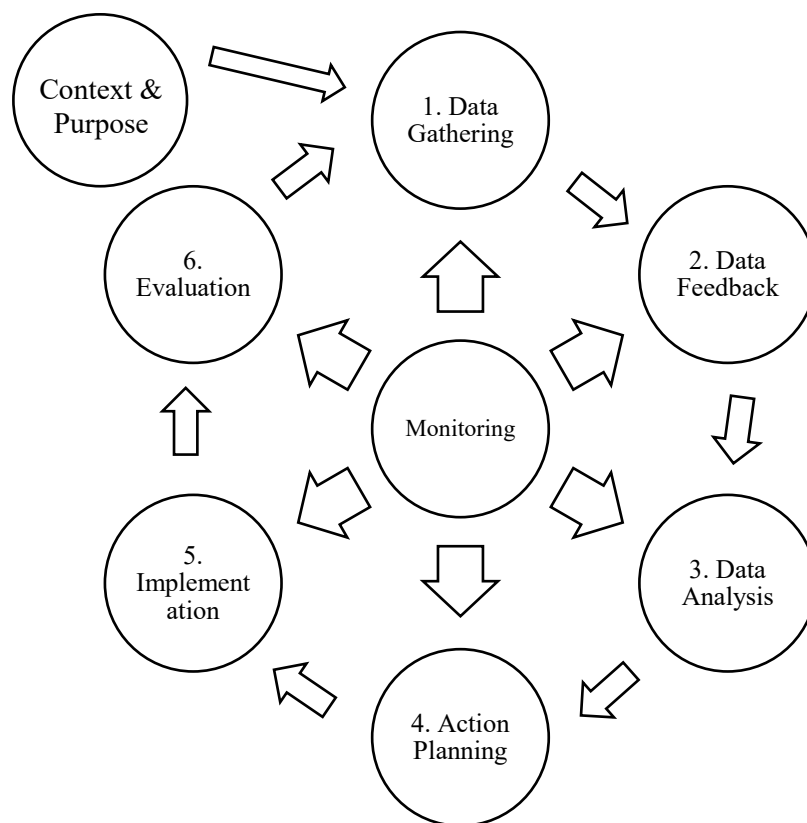
Figure 6 Action research cycle (Coughlan & Coghlan 2002)

The research cycle starts with a separate pre-step to define the rationale for action as well as for research. Once the project necessity and driving forces are specified, then the data is gathered in active daily participation for the organizational processes related to the project. After collecting the data, it must be fed back to the organization to collecting valuable feedback for the data analysis. (Coughlan & Coghlan 2002.) Indeed, in action research, the data analysis is usually done in close collaboration between the researcher and the researched organization (Eriksson & Kovalainen 2011). The collective approach is emerging from the assumption that the clients know their organization best. Data

analysis is followed by action planning, which means allocating the resources and agreeing on the schedule. It can also involve identifying other change management-related factors, such as the type of change and the resistance to change. After finishing the planning, the client is responsible for implementing the planned actions. At the end of the project, the evaluation of the intended and unintended outcomes takes place. The evaluation is the most critical measure for learning. It facilitates improving the actions in the subsequent action research cycles and avoiding the same mistakes. In addition to six main steps, monitoring is a continuously existing action, which purpose is to serve as a method of planning, coordinating, and learning from the action research life cycles. (Coughlan & Coghlan 2002.) All the steps taken in this research are summarized in table 3.

3.2.2 Context and purpose

Over the past recent years, Sievo, the procurement analytics company, has experienced high growth with many new clients. An integral part of their service offering is data warehousing and customer service-oriented activities, which are not automated, standard and scalable in relation to the organization's growth. It has led to a substantial increase in manual consolidation and classification work. The quality of customers' purchasing data is so inconsistent that the product development has challenges to automate all the data transformation work. Currently, all their data transformation activities are located in Finland and the U.S, meaning that the general cost level of the manual workforce is high compared to the global level. Before, the workload has been so low that the disadvantages would have outset the benefits of outsourcing or insourcing. However, the strong growth has made the option possible, and further research was needed. The management of the company requested the researcher to investigate the potential tasks as well as outsourcing and insourcing options more profoundly before making any decisions.

The objective of this research project was to guide the company through the IT Outsourcing decision process shown in Figure 4. The project was named as "Data Enrichment Operating Model Analysis," and the researcher was one of the two project management team members. Figure 7 presents the governance model of the project. There were five major stakeholder groups: steering group, project management, data transformation stakeholder group, outsourcing/insourcing stakeholder group, and other employees. Steering group mainly consisted of the executive team: CEO of Finnish entity, CEO of US entity, CFO, vice president responsible for People & IT, vice president for sales, and vice president for solution implementation. The steering group's role was to provide

guidance, direction, and priorities. Additionally, the steering group was responsible for the final decisions and signoffs. They also guaranteed that the required resources were available.

Besides the researcher, the vice president for the implementation organization was part of the project management. The project management was responsible for planning and managing the day-to-day project activities, ensuring that the objectives are met on time, keeping the steering group updated, facilitating the interviews, conducting the analysis, and producing the action recommendations for the steering group. Data transformation stakeholders were DM Specialists, DM Team Leads from the implementation and ongoing service organizations, Customer Success Manager, implementation Project Manager, and spend module Product Manager. They provided their expertise to define the potential project scope, meaning the possible repetitive data transformation tasks to be outsourced or insourced. Outsourcing/insourcing stakeholders consisted of employees from different parts of the organization: Vice Presidents for Sales and People & IT, CISO, CFO, CEO of US entity, DM Team Leads for the implementation and continuous service organizations, Consolidation Champion, Classification Champion, spend Product Manager, and spend Product Owner. Their role was to offer insights regarding their specific expertise for building the different outsourcing and insourcing scenarios.
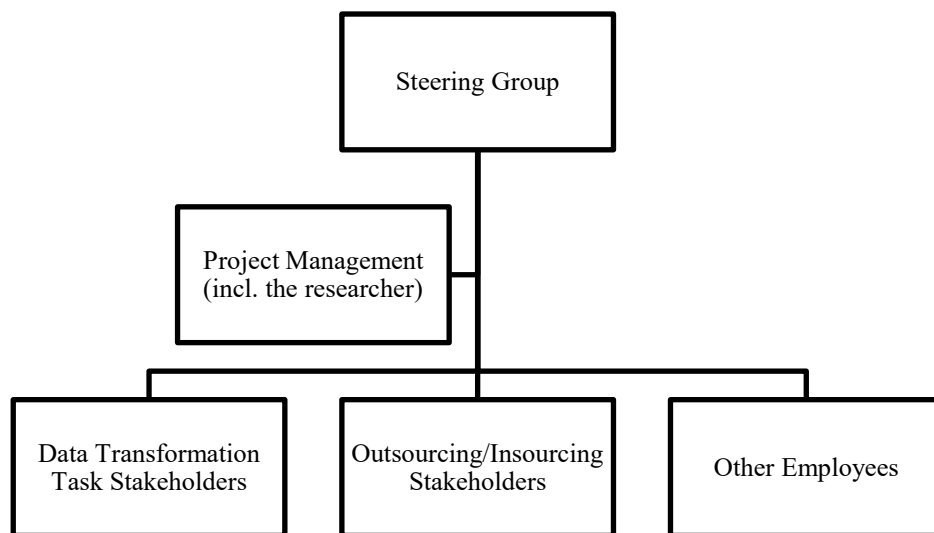

Figure 7 Project governance model

From the academic perspective, the purpose of the study was to showcase the procurement analytics BIaaS organization going through the IS outsourcing research phase. Additionally, the dissertation aimed to provide an in-depth description of the procurement analytics data transformation tasks. Finally, the ultimate objective was to

conduct a direct comparison between data transformation outsourcing and insourcing with respect to potential benefits and costs.

3.2.3 Data gathering and feedback

In action research, data is typically collected through multiple methods serving the research objectives (Eriksson & Kovalainen 2011). The data was gathered by observations, secondary source discovery, stakeholder group interviews, and one-on-one interviews in this research. The data collection took place in multiple cycles. Decision process phases and application of outsourcing stages by Dibbern et al. (2004), Sievo project stages, and data gathering cycles are all presented in Table 2 for clarifying the purpose and outcome of each data gathering round.

Table 2 Decision process phases & outsourcing stages by Dibbern et al. (2004), Sievo project stages, and data gathering cycles

Decision Process Phases, Dibbern et al. (2004)	Application of Out-sourcing Stages, Dibbern et al. (2004)	Sievo Project Stages	Data Gathering Cycles
Why	Determinants	Understanding the root cause	1st data gathering cycle:
	Advantages/disad-vantages	Defining the evaluation criteria	Literature, steering group, and employee feedback
What	Degree of outsourcing	Defining the outsourcing/insourcing scope: the repetitive data transformation tasks	2nd data gathering cycle
	Degree of ownership	Evaluating the consequences of changing the ownership or location	3rd data gathering cycle
Which	Guidelines, procedures, and stakeholders of decision initiation, evaluation, and making	Arriving at the optimal data transformation operating model by following TCT	4th data gathering cycle

.

The aim of the first data gathering round was to understand the root cause and the actual problem to be researched. It included unstructured interviews with the CEO, DM Team Leads, and Data Managers. Moreover, secondary sources such as financial reports and monthly meeting notes were utilized.

The second data collection round aimed to define all the data transformation tasks potentially transferred to an external service provider or nearshore/offshore subsidiary. The second round included using the organization's secondary sources, observations, and stakeholder group interviews. The first stakeholder group interview concerned supplier consolidation validation, and the second one spend classification. Both stakeholder interviews took one hour, and they were held in English.

The third data collection round comprised nine in-depth interviews: eight one-on-one interviews and one one-on-two interview. The objective of the data gathering round was to hear employees from different parts of the organization and obtain a comprehensive overview of the potential consequences of outsourcing and insourcing. All of the interviews took place online via video call, and they were recorded for further analysis. The interviews lasted from 30 min to 120 min depending on the project's direct impact level on the interviewees' responsibility area. For example, the interview with the Sales Lead took 30 min while the interview with the implementation organization's DM Team Lead lasted 120 min. Five out of nine one-on-one interviews were held in English, whereas the rest of the interviews were in Finnish because the interviewees felt that they could express themselves more effectively in Finnish. Due to the complex nature of the research project, all the interviewees received an interview document few days beforehand. The form contained an introduction to the project context and the actual interview questions to orient themselves and be capable of providing rich answers to the interview questions. The interviewees were given the questions from a larger question pool based on their field of expertise. The whole question pool can be found in Appendix 2. All the questions were asked at least from one interviewee, whereas some were asked from multiple employees.

The main objective of the fourth data collection round was to align the qualitative and quantitative results of the study and arrive at the recommendation of the most optimal operating model according to TCT. The fourth data collection cycle consisted of several meetings with the CFO. He was mainly responsible for the quantitative cost analysis of the different outsourcing and insourcing options, whereas the research took care of qualitative metrics of the evaluation criteria. The study also assisted in evaluating the time

allocation between the different data transformation tasks. Those results can be found in Table 5.

Data feedback allows the researched organization to be involved in the data analysis, action planning, and evaluation procedures (Eriksson & Kovalainen 2011). In this research, continuous feedback was enabled by an open communication channel in the organization's intranet. All the organization members had unrestricted access to the project documents, and their opinions were heard regarding the gathered data and the used framework. During the project, two official status updates were held for all the company's employees, facilitating one additional channel for feedback and questions. Sometimes it was challenging to receive input from the data transformation-related employees automatically. In those cases, the employees, such as DM Team Leads, were directly contacted and asked for feedback. In addition to company-wide and individual feedback channels, the steering group meetings served as a gathered data feedback channel. However, due to the limited time frame, the steering group did not often comment on individual data points but instead provided higher-level guidance. All the held interviews and other feedback sessions are marked in Appendix 1.

The feedback approach was especially exploited when the researcher established criteria for evaluating different data transformation tasks within chosen scenarios. Instead of starting from scratch, the researcher collected the most common outsourcing benefits and risks from the academic literature and translated those to evaluation criteria. After that, she asked for feedback from the open internal communication channel dedicated to the project. This approach enabled tailoring the evaluation criteria corresponding to the organization's needs.

3.2.4 Data analysis and action planning

Data gathering took place in several cycles. However, the data analysis was not every time conducted in a similar fashion time due to variating nature of the data and the information needs. During the 1st and 2nd data gathering cycles, the data analysis was very straightforward since the researcher worked with those topics on a daily basis as a DM Specialist and as a participant-observer. Additionally, the CEO, implementation DM Lead, stakeholder group interview participants, and Classification Champion were able to provide a detailed explanation for root causes and data transformation tasks. Moreover, the organization already had some written exiting documentation of the root causes. The research just gathered the information together for power point, word, and excel files.

The most resource-intensive data analysis concerned the 3rd data gathering round. Data analysis started with transcribing the interview recordings. The interviews held in English were transcribed by exploiting Otter.ai text transcription software. To guarantee the transcription quality, the researcher scrutinized the content and corrected possible mistakes and spelling errors. Whilst interviews held in Finnish were manually transcribed and, after that, translated into English by the researcher. After completing the transcribing, the interview content was coded corresponding to *a priori* defined IT outsourcing criteria (table 1). Then, the interview data were utilized to add rich content to the scenario matrices, where different data transformation tasks and outsourcing criteria represented the y and x-axis. After having a qualitative description for each task within each scenario, the researcher gave them a numeric value from 1 to 4 depending on the results obtained from the interviews.

The fourth data gathering round was more collaborative in nature. It was about making sure that the qualitative and quantitative models are aligned with each other. The researcher and CFO decided which factors are considered as financial costs and which factors should be just considered as negative qualitative costs. In the end, the fourth data gathering cycle provided the quantitative cost estimation values for the TCT approach.

The project management always did initial action planning. However, the steering group gave strong direction and guidance for initial action plans. For example, the steering group decided on the potential operating models to be researched. The steering group meetings were often held after finishing a data-gathering cycle. In those meetings, the results of the previous data analysis and future action planning were presented. It facilitated the steering group being up to date, confirm the planned actions, and provide feedback.

3.2.5   Implementation and evaluation

Usually, in action research, the client organization implements the planned action, but in this case, the researcher was responsible for the practical implementation (Coughlan & Coghlan 2002). She guided the organization through the IS research phases: *why*, *what*, and *which*. She organized each meeting, interview, and feedback session. The research managed the project schedule and made sure that every critical aspect was taken care of.

Besides being part of the project management, the researcher also worked as a DM Specialist to enable active involvement in the data transformation work. However, her dual role was informed to the other employees at the beginning of the project to ensure

that everybody is aware of observational data gathering on an ongoing basis. The data collection was executed mainly through formal openings, such as meetings dedicated to the project, guaranteeing consent for the data gathering. Informal openings, such as DM specialists' coffee breaks, were rarely used for data collection. In that setting, the participants were aware of the gathering and the usage of such materials.

The action researcher is an active member of the researched organization; therefore, the ethical issues require special attention (Salmela 2008). Before choosing the researcher for the project, the company's top management had already decided that none of the employees will be laid off regardless of the final results of the research project. The researcher worked in the role that is the potential subject of outsourcing or insourcing. Without guaranteeing employment in the future, there would have been a major conflict of interest in researching the possible outsourcing of her own role. However, even though there is no risk of losing employment due to the research, the results might significantly impact the researcher's everyday working life as a DM Specialist. That is why the researcher needed to pay attention to be neutral throughout the project and not let her individual experiences or personal relationships within the organization create bias to the obtained results.

Evaluation typically concentrates on actions taken during the research process. It includes both the intentional and unintentional consequences within the organization (Eriksson & Kovalainen 2011). The researcher did not have previous experience in conducting the IS outsourcing research phase, which was shown in a way that she was unsure about the best practices. Especially when it comes to the level of accuracy and the number of criteria to be analyzed, the lack of previous project experience was demonstrated. In the end, the final results contain ten criteria concerning five tasks within five scenarios. The interview results decreased the number of scenarios from five to three, but still, it resulted in 150 individual values. Even the steering group was not interested in the results in that detailed level, but they preferred scenario-level aggregated values.

Data transformation outsourcing/insourcing was not only a new topic, but this was also the first time that the researcher used action research methodology. Generally, action research is not a simple approach, but it requires confidence for coping with the uncertainty and capability to be exposed to an organizational change in real-time (Coughlan & Coghlan 2002). The researcher had just started working one month before in the organization, which meant there was no confidence emerging from the organizational expertise. The very experienced other project manager provided his guidance and

knowledge for running the project; however, he was somewhat in the role of a mentor rather than a fellow researcher. The researcher executed everything alone. High-level uncertainty and lack of previous experience coupled with a very tight schedule resulted in very high pressure and potentially impacted the final output quality. However, in the big picture, the researcher managed to finalize the research phase and complete all the required sub-steps defined by Dibbern et al. (2004).

3.2.6   Monitoring and limitations

While the project steering group focuses on the practical outcomes, the research is responsible for monitoring the learning process (Coughlan & Coghlan 2002.) In this action research, the project management, the researcher, and the Vice President for the implementation function had a fixed project monitoring meeting twice a week. Those meetings are not marked in Appendix 1 since those took place regularly, and their function was project monitoring, not data gathering or decision-making. Typically, in those meetings, the project management reviewed the latest actions and findings as well as agreed on the next steps.

Action research as a methodology is prone to certain limitations. It has a tendency to be confused with consulting and that kind of approach is often expected by the client organization, which is financially supporting the research. It is possible that the researcher becomes heavily involved in the problem setting and forgets her obligations of developing general knowledge about the related theories. (Baskerville 1999.) Also, in this case, the case organization was funding this research setting the practical contribution as a primary objective outclassing the academic endeavors as a secondary target. In the end, this conflict was solved in a way that the researcher provided different final products for the academic and practical purposes for contenting both parties.

Table 3 Action research steps

Phase	Description
Context & Purpose	The amount of manual consolidation and classification work has substantially grown. The automation is not capable decrease the workload enough. Finland nor the U.S are cost-effective locations doing the repetitive work. So far, the workload has been so low that the disadvantages would have outset the benefits. However, outsourcing or insourcing could be a feasible option with continuous growth, and further research is required.
1. Data Gathering	There were several data gathering cycles for multiple purposes by using different qualitative methods. The data was gathered from participant observations, secondary sources, one-on-one interviews, and data transformation task stakeholder group interviews.
2. Data Feedback	The researcher facilitated multiple feedback channels; company-wide update sessions, internal open communication channel, full access project documentation, direct feedback requests, and steering group meetings.
3. Data Analysis	The analysis of certain data gathering rounds (1, 2, 4) required collecting already well-analyzed and straightforward information for a common format. On the other hand, data gathering round 3 included transcribing, analyzing, and summarizing unstructured interview data.
4. Action planning	After each data analysis round, the initial action plan was created by the researcher and the rest of the project management. After that, the steering group always gathered, and the latest findings and next planned actions were presented. The steering group provided guidance and direction for the following actions.
5. Implementation	Once the actions were planned, the researcher always implemented all the required actions: meetings, interviews, material creation, and other data gathering.
6. Evaluation	The researcher did not have previous practical experience of data transformation outsourcing/insourcing nor action research. It was shown as high pressure and uncertainty. All the research choices, e.g., the level of detail was not optimal. However, the researcher managed to finalize the research phase and conduct all the required steps.
+ Monitoring	Project management had twice a week fixed meeting for introspective monitoring.

4   RESULTS

The research project aimed to follow Dibbern et al. (2004) *decision process;* therefore, the results are also reported according to those three main steps: "why," "what," and "which."

4.1   Why

4.1.1   Motivation for the outsourcing or insourcing project

Sievo's recurring revenue has grown by an average of 30% per annum during the past five years, and the trend is expected to continue in the upcoming years. In BIaaS organization, it means that the number of clients has steadily grown. Naturally, the client implementation projects occur only once, and overall, there are peak times and calmer periods. However, the total amount of long-term continuous service workload increases every time with a new client. Additionally, the clients have increased expectations of data transformation quality, efficiency, and customer service as a top-end solution. When thinking of manual data transformation, the optimal solution would be technical automation; as the CEO of Sievo states, "We will continue to automate as a big share of this work as possible. This is obviously the best way to reduce workload - better have software than human doing this work." However, at the moment, there is no simple technical solution available for the growing amount of repetitive data transformation work.

Sievo cannot just automate the current data transformation work and create an AI, classifying and consolidating all the procurement transactions. In most cases, the data coming from the clients have poor quality. In fact, one of the DM Specialists states that "In almost 80 to 90% of the cases, they [clients] have very bad data quality and that is the reason why they come to us; otherwise they could just create a Power BI by themselves based on their data warehousing, and they would not need Sievo." Poor data quality in classification is emerging because the clients are recording their transaction data in ERPs in a manner that does not facilitate precise and accurate data points. Instead of having clear data points, the transactions are described with open text fields making systematic approach and algorithm training difficult. The manual work is instead Sievo's competitive advantage rather than a burden that should be just eliminated at any cost. The Product Manager argues, "I think that the human involvement is also part of our competitive

advantage. We do not just rely on these automatic tools, but we rather have the actual expert input, whether this expert would be on the client-side or Sievo's DM Specialist."

Additionally, Sievo typically serves big multinational clients with high expectations. The standard solution does not fit all the clients and their unique needs. Indeed, Classification Champion says, "I think the value that we add to the clients is that we can custom configure all these things, e.g., clients want to organize spend categories in a certain way. The more complex the client organization is, the more custom configurations there can be. If they are very small clients without previous spend analysis expertise, in those cases, they might be fine with the standard approach." The more standardized solution is offered, the less cooperation with the client is required.

So far, Sievo has been tackling the growing amount of repetitive data transformation work by hiring new DM Specialists in both locations US and Finland. Nevertheless, these locations are not cost-effective options for executing such repetitive work. Additionally, one of the company's principles is to offer the same high level of benefits for all the employees, increasing the total cost of full-time equivalent (FTE) DM employees.

4.1.2 Evaluation criteria

For evaluating the benefits and risks emerging when changing the location and ownership of the data transformation tasks, the researcher was required to establish a suitable evaluation criterion for the current situation. The initial criteria were gathered from the academic literature, as shown in Table 1. After collecting feedback from the steering group and other employees, the criteria were modified to correspond to this particular setting and research objectives. This criterion is slightly narrower compared to the one based purely on the literature. Even though the actual cost analysis is part of the real-life project, this research paper focuses on researching the qualitative criteria, which is why this criterion does not include task-level cost estimations.

Moreover, required scalability, need for resource optimization, and sensibility to instability are excluded from the criteria. Maximum scalability and resource optimization are both long-term objectives for the data transformation tasks development rather than just being equally important criteria. Therefore, there was no necessity for analyzing those at a detailed level. Moreover, sensibility to instability is linked to other criteria, which means that it will be analyzed primarily through efficiency and quality requirements.

In this research, all the qualitative criteria are treated with equal importance. Still, in practice, the organization could create a priority order for the criteria based on the most

desired benefits and the most impactful risks. Later in the next subchapters, each criterion is verbally and numerically assessed at task-level accuracy. Numeral values are given from one to four (1=low, 2=medium-low, 3=medium-high, and 4=high). It is essential to understand that there is no one optimal value for each criterion. Table 4 provides a short description of each criterion and demonstrates whether the value should be high or low at the optimal situation.

Table 4 Evaluation criteria

Criteria	Explanation	Optimal Score
Efficiency	Operational efficiency	High
Quality	Quality of data transformation work output	High
Technological development capability	Organization's capability to develop its data transformation tools	High
Positive impact on customer perception	Client's perception of the organization	High
Positive contribution to competitive advantage	Task's ability to positively contribute to organization's competitive position	High
Relative capability position	Organization's capability to conduct the data transformation task in a superior manner compared to competitors or other service providers	High
Potential for opportunism	Risk of becoming overdependent on supplier's ability to conduct the specific data transformation task	Low
Required expertise	Required expertise for conducting the data transformation task (technological, industry, and customer-specific expertise)	Low
Required management capability	Required managerial effort for having the tasks conducted successfully	Low
Required HR capability	Required HR effort for having the tasks conducted successfully	Low
Required Information capability	Required information security effort for having the tasks conducted successfully	Low

For efficiency, quality, and technological development capability, the optimal value is as high as possible. Naturally, the higher the operating efficiency and production quality, the better. High technological development capability instead means that the organization has prerequisites for developing its data transformation capability.

Moreover, the positive impact on customer perception, contribution to competitive advantage, and relative capability position should be as high as possible since the data transformation tasks are an integral part of Sievo's service offering. Positive contribution to competitive advantage reflects the tasks' ability to enhance the organization's competitive position. Instead, relative capability measures the company's unique ability to perform the task in a superior manner compared to competitors and vendors. It would not make sense pursue relative capability position or competitive advantage in repetitive work in some other cases. However, data transformation tasks are generally perceived as the key competitive differentiators, and therefore it is desirable to improve those capabilities.

The potential for opportunism measures the risk of becoming overdependent on the external service provider, which applies only to outsourcing. The lower the potential for opportunism, the smaller the likelihood of becoming locked to a single vendor. Required expertise, required management capability, required HR capability, and required information security capability refer to different needed abilities for running the tasks successfully. Obtaining a capability always requires some sort of investment. Therefore, having a low value in those criteria indicates that fewer resources are needed to allocate for training, management, HR, and information security.

4.2   What

After understanding the root causes for the issue, the aim is to comprehend in detail what would be the optimal tasks to be included in the project scope and how the tasks are currently conducted. Indeed, the degree to which the process can be codified will also influence the ease of transferring it to another entity (McIvor, 2008). Additionally, it is vital to acknowledge the major implications of changing the location and the ownership of the data transformation tasks before drawing any conclusions, as Dibbern et al. (2004) suggest. Figure 8 demonstrates the different operating model scenarios that were considered as an option. For evaluating the benefits and risks, first, the results describe the cur-
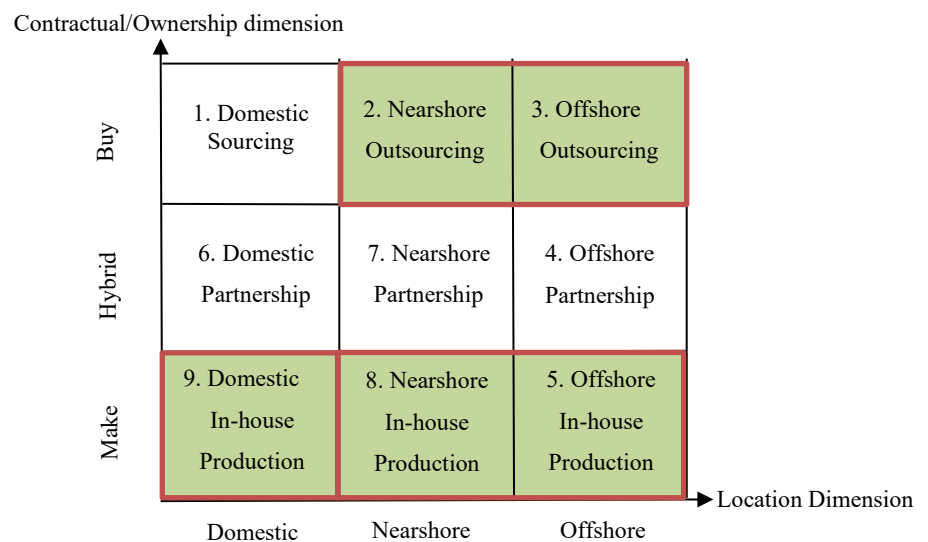
Figure 8 Selected outsourcing and insourcing scenarios

rent state, which corresponds to scenario 9, domestic inhouse production. After that, the results demonstrate the potential consequences of nearshore and offshore outsourcing (scenarios 2 & 3). Finally, the chapter describes the potential benefits and risks related to nearshore and offshore inhouse production (scenarios 5 & 8). The results illustrate that the difference between nearshore or offshore is not prevalent decision driver at the research phase of the IS outsourcing process. Therefore, the final results are aggregated into three alternative scenarios domestic in-house production (Table 6), outsourcing (Table 7), and insourcing/establishing a new entity abroad (Table 8).

4.2.1 Current state: domestic in-house production

In the first steering group meeting, it was agreed to limit the project scope to repetitive, mass-production type of data transformation tasks. Hypothetically those tasks could provide the highest cost savings, and transferring those to another entity would be relatively easy compared to other tasks. Those are entry-level tasks, which do not require years of previous experience or extensive knowledge transfer to be executed successfully. The management did not see it as the most feasible option to consider the more demanding and complex expert work to be potentially outsourced or insourced. Data transformation is a high-level term for all the tasks. Still, the objective is to understand in-depth what type of data transformation tasks DM Specialists do in both implementation and continuous service functions and which of them should be included in the project scope. In Sievo, data transformation activities can be divided into three main categories spend classification, spend classification validation, and supplier consolidation validation. Moreover, reclassification and reconsolidation are considered separate correcting tasks. All these tasks can still be broken down into more detailed sub-steps. A detailed description of all the sub-steps can be found in Appendix 3. The main principles of the data transformation tasks are the same in both implementation and continuous service functions, but the nature of the work is still slightly different. The total workload of these pure repetitive data transformation tasks is notably higher in the implementation phase compared to continuous service. Table 5 provides short descriptions of each repetitive task and showcases their resource intensity. Furthermore, in this chapter, the tasks are evaluated against the criteria defined in the previous chapter. The subchapters provide a qualitative description of each value, and Table 6 demonstrates the numerical evaluation for each task.

4.2.1.1 Spend classification

Spend classification is about harmonizing all purchasing transactions to a client-specific cost hierarchy. In the implementation phase, the purchasing transactions coming from the clients are profiled, meaning that the data is clustered based on the information that it contains. That information can be, for example, purchase order (PO) numbers, material numbers, account information, vendor information, free text POs, cost centers, and invoice descriptions. (Sievo 2021)

Table 5 Aggregated list of repetitive data transformation tasks

Aggregated Task List	Description	Cus-tomer-facing task	Resource intensity (% of total work-ing time/full-time equivalent em-ployee (FTE))		
			Continuous service	*Implementa-tion*	*Other Functions*
1. Spend Clas-sification	Classification is about har-monizing all purchasing transactions into a single taxonomy. Analyzing data points and manually choose the correct spend category based on the available data points.	X	11%/1.4FTE	35%/4.7FTE	-/0.3TE
2. Internal classification validation	Usage of different validation methods for guaranteeing the accuracy of the made classi-fication decisions.		<1%/0.1FTE	15%*/2FTE	-/0.6FTE
3. Consolida-tion validation	Validating the supplier con-solidation decisions made by AI are correct. In case of er-ror, update or create a new supplier to the system.		4%/0.5FTE	10%/1.3FTE	-/0.1FTE
4. Repetitive reclassifica-tion	Correcting incorrect spend classifications based on the ad-hoc reclassification re-quests coming from the cli-ents.	X	6%/0.8FTE	4%/0.5FTE	-/0.1FTE
5. Repetitive reconsolida-tion	Correcting the incorrect sup-plier consolidations based on the ad-hoc reparenting re-quests coming from the cli-ents.	X	3%/0.4FTE	-	-
Total			24%/3.1FTE	64%/8.6FTE	-/1FTE

*Includes also external classification validation hours

Table 6 The current state of repetitive data transformation tasks

Task	Efficiency	Quality	Tech. dev. Capability	Pos. impact on Cus. Perception	Pos. Contribution to Competitive Advantage	Relative Capability Position	Req. Expertise	Req. Management Capability	Req. HR capability	Req. Inf. Sec. Capability
1. Implementation: Spend Classification	3	2	3	4	3	4	3	2	2	3
2: Implementation: Internal Classification Validation	3	2	3	2	2	1	2	2	2	3
3. Implementation: Repetitive Reclassification	3	2	3	3	3	3	2	2	2	3
4. Implementation: Supplier Consolidation Validation	3	1	3	1	2	1	2	2	2	3
5. Continuous Service: Spend Classifications	3	2	3	3	3	3	3	2	2	3
6. Continuous Service: Repetitive Reclassification	2	2	3	3	2	3	2	2	2	3
7. Continuous Service: Consolidation Validation	3	1	3	1	2	1	2	2	2	3
8. Continuous Service: Repetitive Reconsolidation	2	2	3	3	2	2	2	2	2	3

Additionally, experts of Sievo build in cooperation with the client a taxonomy of different spend classes. The taxonomy is created corresponding to the clients' cost structure and data quality. In the implementation phase, manual, repetitive work is required for classifying all the historical data from zero until 90% classification coverage of total spend. DM Specialists are responsible for ensuring continuous 95% classification coverage for the new monthly or weekly incoming purchasing data on the continuous service function.

Sievo utilizes classification Artificial Intelligence (AI) that is supposed to learn from the existing classification decisions. AI recognizes from the already made classification decision the specific combination of data points, e.g., account information and vendor information. In the future, the AI applies the same decision for the new incoming transactions with the same data points. The current process is already highly optimized. Indeed, one of the interviewees stated that "If we have clear mapping, if we have clear taxonomy and very reliable data points, it's very efficient already, just like a few hours of work." However, in most cases, Sievo receives insufficient quality data from the clients, increasing the amount of manual work required. Suppose the data points describing the new incoming transaction data are, e.g., free text fields or vague keywords. In that case, the AI has difficulties learning from the historical decisions and identify the correct classification. In those cases, DM Classification Champion says, "We are also doing the enrichment work. For example, if the vendor does not have any usable information, we need to google it. It is what people often refer to as manual. If I had to google like 100 vendors for what they are doing. Sometimes it's easy, but sometimes it's not that straightforward." Currently, Sievo is not prioritizing that DM Specialists would develop towards industry experts, meaning that they would be specialized to, e.g., direct materials manufacturing companies. Poor data quality also requires strong cooperation with the client. According to continuous service DM Lead, "[Poor data quality] requires us to go through a significant number of different kinds of rules on how the client wants the data to be classified and special conditions." In classification work, efficiency is not prioritized over quality. Implementation DM Lead remarks that "When we personally know the clients, it creates mental pressure to make sure that the classifications are correct, which leads DMs sometimes spending too much time for individual decisions."

The importance of human involvement in classification work is emerging from the fact that classification quality is a critical part of Sievo's offering. The classification quality can be perceived in two different ways:

1. First-time quality

2. Continuous quality

According to the CEO of Sievo Inc, "First-time quality means that if somebody gives me information of the data, which is wrong. Can we correct it? Can we really make that correction?" According to implementation DM Lead, "when clients notice incorrect classifications on the analytics dashboard, they either directly send an email with correction request or a ticket through Sievo's customer support service portal. When it comes to reclassification, the quality is naturally high since those classifications are based on direct client feedback. On the other hand, it is also possible that client feedback is not either completely accurate. According to current service level agreements, those change requests must be taken care of approximately within five working days. The more challenging part of first-time quality is the classification decisions made based on DM Specialists' research. The classification quality is also heavily linked to the quality of incoming purchasing data. If the data does not include useful data points, the employees need to understand the nature of the transaction based on the available information online. DM Classification Champion concludes, "[Using googling for data enrichment] means that we need to compromise on the accuracy. Because one vendor might do many different things, and maybe they supplied completely different things to the client [than found online], but we just don't have the data for it."

Continuous quality instead measures how well AI can learn from the executed decisions and apply the findings for the new incoming data on a monthly basis. CEO of Sievo Inc. says, "[Continuous quality is a] competence that we definitely need to build more. Software is a key to that. But it's also people who are maintaining and monitoring the software, that it's actually doing proper things, developing those algorithms, and developing those ways of working forward." Classification decisions train the algorithm. Therefore classification quality has far-reaching consequences. According to the implementation DM Team Lead, "In general, it [classification] is given more attention than the other repetitive tasks in the implementation phase. Quality is preferred over efficiency, but it is an endless debate in the implementation organization when to prioritize time constraints over quality. In the end, if the classification quality is poor at the project phase, it fires back, and we need to increase it later stages anyways." Therefore, spend classification solely is the most resource-intensive task out of all the repetitive data transformation work.

Despite the importance of classification quality, currently, Sievo does not strictly follow classification decision quality metrics. The best indicator for that is the client

feedback in the implementation phase as well as in continuous service. Clients pay active attention to the classification quality. According to the implementation DM Lead, "We often do not get negative feedback of our total offering, but we get negative feedback of our classification quality." Despite the few negative feedback concerning the classification quality, the Head of Sales says, "Sievo has an overall positive reputation as a spend analytics SaaS company. Sievo is considered a market leader in data extraction and transformation activities in Europe." The importance of classification is very present in sales. According to the Head of Sales, "PoC classification work is the way of convincing potential clients and crucial factor impacting directly to client's purchasing decision."

DM Specialist is an entry-level position in the organization. Indeed, the implementation DM Lead says that "People usually start with little practical knowledge of data management, but people can take care of the tasks independently after the training." Similarly, the continuous DM Lead states, "I think our software is not necessarily the super easiest when we're talking about classification and consolidation. But it is something that is with proper training understandable. So, I would say that a minimum requirement is logical thinking and being somebody who is also analytical and tries to understand software or has software experience, but just facing, not developing." Additionally, some of the tasks are customer phasing; therefore, the continuous DM Lead adds. "The person would also need to have a bit of customer-facing experience like sitting there being able to guide and steer customer discussions or needs and have a proper command of English." In the end, the role-specific expertise grows over time when DM Specialist learns the tools and the client-specific details. In continuous service, according to the DM Lead, "Some customers require more knowledge, because of different rules because of data quality, to be properly handled on a monthly basis." Additionally, the documentation of client-specific information is not at a sufficient level. Classification Champion mentions, "We are not able to provide a taxonomy explanation in most of the cases. In real life, I realized that the clients do not even have any explanation on the taxonomy by themselves". It means that sometimes even an experienced DM Specialist cannot frictionlessly start taking care of a new client if she/he is not aware of client-specific customization. The implementation DM Lead also underlined the nature of DM Specialist, "DM Specialist usually stays at her/his current role only for a year and then moves forward, which means that there are no people with years of expertise conducting the classification tasks."

Classification is one of the main drivers for the clients choosing Sievo's solution over the competitors. According to implementation DM Lead, "Classification has a

significant contribution to our competitive advantage. We do it very differently from many of our competitors, and it is a significant part of our service provision. The usage of a combination of multiple data points has been a new thing for several of our clients when they implement Sievo. Clients can use, e.g., material group, account, and vendor information for having exactly the correct classification. Additionally, even though the poor incoming data quality impacts the classification efficiency and accuracy, the overall level is still superior compared to the competitor. CEO of US entity says, "Our classification engine and the approach that we take to get 95 to 98% of the spend classified in a relatively accurate manner and very fast, is a competitive advantage that we have over the others." Despite the high importance of classification for Sievo's competitive advantage, the actual relative capability position or uniqueness does not stem from the manual, repetitive classification work. Implementation DM Team Lead argues that "Unique for us is the fact that we do not do pure AI, but also not pure manual work. We can use all the data points, and it reflects how we build classification taxonomy in the classification tool."

Regarding technological tools used for classification work, there are three different interfaces for executing them – directly in the database by using SQL, in the classification web tool, on the dashboards using the in-analytics classification feature. Database classification is mainly used in mass classifications by importing reclassification excel files in a particular format directly to Sievo's database. This work is conducted only by experienced DM Specialists, and clients never have direct access to Sievo's database. The second option, the classification web tool, is the most popular tool for classifications. Some of the clients' super-users might have access to it, but mainly it is solely used by the employees of Sievo. The third tool is recently launched new feature, enabling classification decisions and suggestions in the analytics dashboards. The idea is to empower the end-user and increase the classification quality based on their feedback. Typically, in the implementation phase and the monthly data classification, the web tool is the primary tool, whereas reclassification requests occur in analytics. DM Classification Champion says, "I see that classification tool is a very sustainable and stable way to create that decision tree for the data." However, when it comes to technological maturity, there is still room for technological improvement. Classification Champion points out that "Product function is still working a lot on the tools. So, that new feature is coming on a weekly or monthly basis. When we want to transfer something, it needs to be stable enough." As seen before, the poor quality of incoming data is one of the main drivers for human

involvement. Despite the workload caused by the poor data quality, currently, there is no development project in progress. According to DM Classification Champion, "I can see that we're still very far away from really increasing the efficiency of any data classification. The reason is that the text-based classification still is not on the priority list in product development right now. However, this manual data enrichment is not a feasible solution in the long-term." In conclusion, the data transformation tools are still under continuous development, and they are also changing in the upcoming years.

4.2.1.2 Spend classification validation

After implementing the Sievo solution and completing the classifying 90% of the total spend, the classification decisions must still be validated. Some of the validation work is done externally with the client, and some work is done only internally. DM Classification Champion says, "The most value-added validation we do with the client, but the reality is that the client cannot be involved in validating all of the spend because we don't have the resources, they don't have the resources." There are different types of validation methods: category validation, supplier split validation, material split validation, and category depth validation. Their main idea is to ensure that the particular service/material/supplier can be found under the correct category in the dashboards, meaning that the transactions are classified correctly. The more detailed descriptions of each classification validation method can be found in Appendix 3. Classification validation is not a fixed part of the continuous service, but it instead takes place after implementation or new data source integration; therefore, continuous service DM Specialists rarely execute classification validations.

Classification validation work is about researching proof from the internet that the transaction is classified correctly. DM Classification Champion describes, "When we do validation, we also need to research what kind of industry and what kind of category that vendor is doing and how it is presented in reporting for the clients. It's very difficult to measure the efficiency of that research work." Classification and classification validation tasks are heavily linked to each other. According to implementation DM Lead, "In most cases, the same person does classification, internal classification validation, and participates in the client workshop. Usually, the person learns things from the actual classification for the validation work, making it more efficient." Often one person is in charge of validating all the classification categories, which means that current DM Specialists do not necessarily develop particular category expertise. Implementation DM Lead also

points out, "Especially indirect spend categories are easy to validate, and typically other DM Specialist can effortlessly jump in and assist with those classification validations." On the other hand, especially with the direct categories, classification validation can be more demanding. DM Classification Champion points out that "Implementation organization can give you an estimate that we can research like 150 to 200 vendors per day, but it doesn't mean that it will always be the case, because sometimes it's very fast, sometimes it's not. And if we want to make our process more standardized, more automatic, we need to find a way to, to take that uncertainty out, like we need to find a way to enrich your data more systematically."

The same difficulties apply to the quality of classification validation than pure classification work to a certain extent. Again, classification validation quality is related to the quality of incoming purchasing data. DM classification champion states, "We need to do a lot of research to dig down into [classification quality]. What does this vendor actually do? If they offer different types of stuff, what is the probability they would supply one product rather than another one." Currently, strict time constraint limits the efforts directed to the classification validation work. Additionally, the validation quality is heavily linked to the level at which DM Specialists conduct the validation work. Implementation DM Leas says, "If the spend taxonomy consists of four levels, we usually validate the second or third level since we do not have more resources. If the second or third level contains 100 categories, the deepest [fourth] might have 500 categories. It means that if the spend was validated on the deepest level, it would multiply the used time by five. On the other hand, when going to the client meetings, the project managers would prefer having the validation done on the deepest level." Having more resources for the classification validation work would directly increase the validation quality. Additionally, it is essential to notice that there are no metrics for measuring classification validation. DM Classification Champion argues that "[Classification validation] is also one thing that I really want us to measure. Like somehow have marked if the transaction is actually validated. So that we could have a percentage of the total spend classified; for example, 75% of the spend is validated. It could become a key performance indicator (KPI)."

Internal classification validation is the last step in the implementation phase before showing the data to the client. The Classification Champion explains the importance of the tasks as following, "It's very important that we at least get like 70% of classifications correctly. It would be very, very embarrassing if we go into the workshop with a client, and then most of the spend is wrong." Internal classification validation as a task

guarantees that the classification accuracy is sufficient and the data is ready to be shown for the clients. In that sense, it can positively impact the customer impression. However, the different validation methods are quite routinary, and at the moment, the company due not have any unique capability for executing them. Additionally, internal validation tasks themselves do not necessarily contribute directly to Sievo's competitive advantage. Implementation DM Lead states, "Clients are not even aware that we conduct this task. It is rather internal quality check." It is essential to conduct internal classification validation as part of the implementation project, but it is not a crucial mean for standing out from the competitors.

Classification validation necessarily does not require years of previous expertise, similarly to classification, but it is still a slightly demanding task. It is challenging if the client is operating within a previously unknown industry to the employee, and they do not provide any precise mapping of its materials. DM Classification Champion says, "So imagine that a DM Specialist needs to research about mechanical parts used in the glass industry. It requires a lot from the employee. [She/he] needs to understand the taxonomy and then how these different suppliers or vendors or materials fit into place. And yeah, it is not easy. It is manual but not easy at all. And then when it comes with a time constraint, it creates a lot of frustration for the people who do the work because, like, it is not something that you can just google like, it needs a lot of thought involved."

Regarding the technological development of classification validation work, some efforts are already made, e.g., classification depth reports in the analytics dashboard. However, there is still room for improving the current classification validation tools. DM Classification Champion suggests that "If product development could come up with functionality that could separate, like validated vendor spend from the rest, and then share that knowledge among the clients, because there are some very big suppliers, but also local suppliers that are very specific to some clients. I will say that I would just validate this spend with the client, rather than using the information on the internet or any other third-party enrichment because they're not always accurate." It means that the core of the classification validation is still the information sharing between Sievo and clients rather than internal classification validation, on which this research project is mainly focusing. Indeed, DM Classification Champion summarizes, "The only one who has the authority to say that is incorrect, is the client. How confident can we be saying that it is 100% accurate? So, instead of focusing on increasing our resources to validate our classifications, we need to come up with a way to work more with a client." Creating a harmonized

validated global repository for all the indirect spend could be theoretically possible but very difficult. DM Classification Champion says that "It's very challenging; every client has different taxonomy. So even when you validate, it might be that technically two similar categories have two completely different names." Consequently, Sievo should start offering standard taxonomy for its clients, which is opposite to Sievo's current service provision.

4.2.1.3 Supplier consolidation validation

Currently, supplier consolidation is done by AI, which enriches the incoming vendors with additional organization structure-related data points offered by an external data provider. However, the automated supplier consolidation results are not very high-quality, and those must be manually validated. The validation is mainly done by looking at the consolidation results and judging whether it seems correct. If it seems incorrect, a DM Specialist usually googles the vendor and tries to figure out the correct mother company and subsidiary relation. Then the DM Specialist either chooses the right normalized supplier from the system or creates an entirely new one.

Since there are limited resources directed to the consolidation validation, the quality is directly relational to the total time spent on the task. According to DM Team Lead, "In the implementation phase, we are checking 90-95% of the total spend, which is around the top 1000 vendors. However, usually, 95% of total spend covers actually only 10-15% of the vendors." One client has thousands of suppliers in their database, and naturally, all of them cannot be validated. That is why Sievo offers reconsolidation service as well. Clients can submit reconsolidation requests through Sievo's ticket system, and DM Specialists execute the requests according to agreed SLAs. According to the Product Owner, "I think you cannot really automate reconsolidation requests. You need to take out one request at a time and then do the requested changes. So that's very manual repetitive work." Indeed, the consolidation validation work is very monotonic in nature. Implementation DM Team Lead argues that "It is impossible to maintain high efficiency of consolidation validation over a long time period."

When a system contains a certain amount of unvalidated suppliers, it also means that the database contains some errors. Even though most of the clients can tolerate the low consolidation quality, it still is not an indifferent task. Implementation DM Team Lead underlines, "Consolidation quality has an impact on our classification efforts through AI and therefore indirect impact on our unique capabilities." Many classification

improvement ideas are dependent on the consolidation quality. Implementation DM Team Lead adds, "Even though consolidations are repetitive manual work if they go wrong, it can break many things in the different modules of the software." This is why Sievo should still aim towards better consolidation quality.

When it comes to competitive advantage, supplier consolidation does not have equal importance to the classification. Consolidation validation is not necessarily differentiating factor in the eyes of the clients. Implementation DM Leads mentions, "Consolidation validation has no notable contribution to competitive advantage. In the clients' historical data, I have seen some data normalizations, meaning that this is rather a routine task of the procurement analytics."

Consolidation task is straightforward to conduct, and it does not require any previous expertise. The only prerequisite is the capability to use google. The task execution does not require direct communication with the client, and the results are not client-specific. Implementation DM Team Lead says, "We could take all our different consolidation projects to one excel file and validate all of them together." This is indeed the direction where the product development is taking the manual consolidation work. When it comes to technological development needs, the central supplier repository is a work in progress. The idea is that all the vendors from each Sievo's clients would be consolidated in one central repository instead of doing the work separately for each client. According to the implementation DM Team Lead, "In the future, this new technological solution will enable us to validate each supplier only once, facilitating better scalability of the current service offering." Besides the global repository solution, the objective is to create also more user-friendly interface making the consolidation validation work more efficient.

4.2.1.4   *Other relevant factors concerning the current state*

The previous subchapters described each task in detail and their unique qualities. Moreover, in the interviews, few factors emerged that are still relevant but cannot be given individual values in the task levels. Those three factors are required HR capability, required management capability, and required information security capability.

Sievo has been nurturing its company culture over the years. Head of People & IT describes the culture as following, "We've been focusing a lot on the on the culture and it's been this kind of one Sievo, we're all equal, and we're all part of a family and everything like we care about each other and help each other, etc." It can be expected that the strong company culture has also been one of the significant positive contributors to the

company's high growth. It is also shown by offering relatively good fringe benefits for all employees no matter at which level of the organization they are working. Head of People & IT mentions, "In the US, all the people we hire, we bring them to Finland every year." Currently, the company employs people with a long-term scope. Head of People & IT describes it as "a career playground, and you can almost become whatever you want." It is aligned with the implementation DM Lead statement, "Currently, we hire people not only for DM tasks, but we look at their overall suitability to Sievo in the long-term and therefore DM employees are overeducated to repetitive data transformation tasks." It means that current highly educates employees' capabilities do not entirely correspond to the tasks' easiness, creating time-to-time frustration and lack of motivation. However, the organization has a strong do-it-yourself mentality. Head of People & IT says, "Typically, we have had quite a strong attitude to do it ourselves everything possible."

Sievo is managing massive amounts of clients' confidential business data. Annually, the company processes the spend data approximately for the worth of hundreds of billions of euros. It means that information security plays a significant role when considering Sievo's service provision and data transformation activities. The company follows fundamental principles of information security, which are confidentiality, integrity, and availability. In BIaaS business, Chief Information Security Officer (CISO) describes the confidentiality as following, "In this context, it means, that only those employees, who need the procurement data for executing the given task, can access it." Currently, DM Specialists have access to multiple data points since the data transformation work can be done more efficiently the more accurate data points are available. Additionally, DM Specialists do not only conduct data transformation tasks, but they might execute other tasks such as small configurations in the database. According to CISO, data integrity means that "Data is tamperproof, that the classification and consolidation results are the only thing that the DM specialists can modify and not, e.g., vendor information." The third principle is availability, which is strongly related to SLAs agreed with the clients. CISO describes, "Data integrity means that the clients have access to their data, we do not lose it, and the software is not down. We have SLAs, which obligate, e.g., certain uptime and classification coverage." In addition to these fundamental principles, the organization uses a large set of other measures to enhance its information security and compliance, such as IT auditing, telecommunication encryption, General Data Protection Regulation (GDPR) compliance, and certifications. Sievo cannot wholly neglect data privacy

questions since the data coming from the clients' ERPs can sometimes include sensitive personal data in the form of reimbursement of expenses.

Currently, the required management capability does not significantly variate between different data transformation tasks, but in nature, it is rather about higher-level coordination. It was also evident in the interviews that DM Team Leads instead talked about the management capability at a more general level than at the task level. Even in the entry-level DM Specialist role, employees are very independently executing their given tasks. CEO of Sievo Inc. describes the way of working as following, "Sievo culture today is as everything is self-service no matter what it is." It means that the employees have ownership of their own doing and learning. Since new people are continuously entering the teams, the DM teams are responsible for training the latest resources, and its coordination often falls to the team leads. Head of People and IT describes it as "Especially in the implementation function, training is a continuous thing that happens. There is all the time new people coming in, and the people are moving to different positions. It is pretty automatic. However, Sievo starts to be such a big company that we should have specific people dedicated to that, especially for the roles where we keep getting a lot of people, e.g., data management." People are given as much freedom as possible. In the implementation phase, managerial tasks are emerging due to the project-based way of working. According to implementation DM Lead, "No traditional corporate management, my role is to ensure that the objectives are met in the given time frame. Data transformation tasks need to be done in a specific order, which requires time management and coordination." Similarly, in the continuous service organization, during the participant observations, it was shown that the continuous DM Team Lead rather allocates his time to hiring, HR type of work, process improvement, and team coordination than monitoring the working in the detailed level. However, when the number of DM Specialists grows, also the required management capability increases. Indeed, the continuous DM Team Lead argues, "In my current position, I would actually need to outsource some of my tasks already. Now, some of the tasks that I have, I can't really follow up."

4.2.2 Changing the degree of ownership: nearshore and offshore outsourcing

As seen before in this paper, outsourcing is proven to be an efficient way of reducing IT costs. However, there are a large set of other factors that can increase or decrease as a result of outsourcing. At the beginning of the research process, the management decided that they would like to research separately nearshore and offshore outsourcing scenarios.

However, in the internal interviews, the vendor's location did not emerge as a prevalent factor. In Table 7, different data transformation tasks are similarly evaluated against the criteria than the current state. Color coding in the evaluation table demonstrates if the value is more desirable (green), the same (no color), or less desirable (red) compared to the current state.

4.2.2.1 Efficiency

Often, outsourcing is expected to create some cost savings because of the lower unit costs. According to economies of specialization, vendor firms should provide some efficiency improvements arising from the accumulated knowledge. (Chang & Gurbaxani 2012.) However, there is only little room for improvement on individual employee level when it comes to classification efficiency. The Product Owner mentions that there is even a risk of decreasing the data transformation task efficiency. He states, "I suppose the speed is one asset that we have, especially on the PoCs, and things like that. So, we do get the data classified and cleansed and consolidated in a concise time frame." The Product Manager continues, "I don't know how much we can rely on the third-party provider that they could, e.g., execute PoC in a day."

In the end, it is limited scale, how much an individual employee can impact on the production efficiency, as DM Specialist concludes, "I will say that the efficiency depends a lot on the quality of created spend taxonomy and client data." Continuous service DM Lead agrees, "Let's take an example telecommunications specific person who knows the ins and outs of any kind of equipment materials that telco companies purchase, that person will still have a hard time if the data quality is poor." The only efficiency increasing factor according to implementation DM Lead is that "When buying the classifications from a supplier, it would maybe decrease DM Specialists' moral pressure to make correct decisions leading to increased efficiency." Classification validation efficiency instead could slightly increase instead if the third-party had category expertise. Implementation DM Lead says, "Category specialization could offer some economies of scale especially in indirect spend classifications, e.g., marketing, legal, and HR classes. Direct spend categories instead are quite client-specific." The efficiency of consolidation validation work is restricted due to its monotonic nature, and it is hard to maintain high productivity over several hours.

Table 7 Outsourcing the repetitive data transformation tasks

Task	Effi-ciency	Quality	Tech. dev. Ca-pability	Pos. Im-pact on Customer Perception	Pos. Contri-bution to Competitive Advantage	Relative Capability Position	Potential for Opportun-ism*	Required Exper-tise	Required Manage-ment Capa-bility	Required HR capa-bility	Required Inf. Sec. Capabil-ity
1. Implementation: Spend Classification	3	3	2	3	4	3	4	3	2	1	4
2: Implementation: Internal Classification Validation	3	3	2	1	2	2	3	1	2	1	4
3. Implementation: Repeti-tive Reclassification	3	3	2	2	3	3	4	1	2	1	4
4. Implementation: Supplier Consolidation Validation	3	2	2	1	2	2	3	1	2	1	4
5. Continuous Service: Spend Classifications	3	3	2	2	3	3	4	3	2	1	4
6. Continuous Service: Re-petitive Reclassification	2	3	2	2	3	3	4	2	2	1	4
7. Continuous Service: Con-solidation Validation	3	2	2	1	2	2	3	1	2	1	4
8. Continuous Service: Re-petitive Reconsolidation	2	3	2	2	2	2	3	1	2	1	4

Outsourcing could offer an opportunity to organize the resourcing differently. Implementation DM Lead states, "In case of buying the consolidation validation from an external service provider, the efficiency-related questions would be managed differently, e.g., a third-party could have multiple people doing the consolidation validation in shorter shifts guaranteeing higher efficiency."

If not thinking about only the task execution on a detailed level but at the higher operational level, outsourcing could still increase operational efficiency. Continuous service DM Lead states that "If we're thinking about an external service provider, we might be able to increase the DM capability significantly faster than if we could in-house, it's simply because there are different training methods, there are different operations already set up, and they just can assign people from their pool of recruitments. So, from the operating efficiency point of view, we might actually also have a short-term dip in operational efficiency. Still, outsourcing could probably offer significant leverage to increase efficiency based on how much manpower we need. "

4.2.2.2   Quality

Classification quality would increase if the outsourcing partner had employees with industry-specific knowledge. Classification Champion mentions, "It takes a few monthly processes of continuous service, so that you know, how each material is used. We simply do not have the expertise and time to understand how a product or machine is produced. If the external service provider had industry-specific expertise, it might be easier for them to point out very fast where the transaction should fall, decreasing the quality discrepancy."

When it comes to classification validation, if outsourcing would enable us to direct more resources to it, some new additional practices could be launched to increase the accuracy of classification decisions. Implementation DM Lead points out that "With additional classification validation resources, we could do different sort of more complex category validations, e.g., comparing PoCs and full implementations." However, in the big picture, Classification Champion sees more value in external validation done in the cooperation with the client rather than investing significantly more in the internal classification validations. She says, "We are not the one who actually uses the data. So, we might be able to tell /[validate] some, but it can never be 100% accurate. I will say that classification validation is not something that can be automatically outsourced or something because we need the client's input to increase the accuracy." Also, in the case of

fully outsourcing classification and classification validation work, Sievo should have better means for defining the actual classification accuracy. Classification Champion argues that "I would challenge that, whether we should call [pure internal validation] accurate. We should come up with a definition of classification accuracy that needs to be approved by the client. That is just something that I think we haven't been able to agree on a common ground; what do we mean by classification accuracy."

When it comes to consolidation validation quality, there is some room for improvement, and outsourcing could not drastically decrease the current consolidation quality. Implementation DM Team Leads says that "If the outsourcing would allow us to allocate more hours to consolidation work, we could increase the overall supplier consolidation accuracy." The potential quality advancements emerging from the additional resources are heavily dependent on cost-efficiency. Moreover, increasing the overall classification and consolidation quality could decrease the amount of reclassification and reconsolidation requests.

Also, in case of having more resources for feedback implementation, meaning reclassification and reconsolidation work, instead of just executing the changes based on the client feedback, we could analyze the situation case by case and use our expertise. CEO of Sievo Inc. suggests, "If we have more time to make those [reclassification/reconsolidation requests], then we could actually do those more properly. Today, we don't have time. We don't have resources. So, it's kind of if somebody tells you that the supplier needs to go into that category, the easiest route is to say, 'Yes, we'll move it there.' Instead of telling the client that, 'no, it should not go there, you should reconsider.' So, if we would have more time on our hands on that, there could be actually more time to investigating those things and then with that, providing better quality."

4.2.2.3   *Technological development capability*

Overall, the data transformation tasks are not a very standardized process yet, making their outsourcing more difficult. Continuous development of Sievo's tools is raising concerns regarding the feasibility of outsourcing all the DM tasks to an external service provider. Classification Champion argues, "Even we [internally] are not up to date with all the new features, so how we are supposed to transfer anything to a third-party." If the external party used Sievo's tools, it would require setting up training regularly, which would demand some additional resources from Sievo. The complexity of the current data transformation tools would set additional requirements for the nature of the partnership.

Product Owner says, "The tools that that we have currently in place are quite complex. So, if we are thinking about outsourcing, then I suppose we are considering ten people that would be having a long relationship with us."

Some non-repetitive DM work is very database heavy; however, the pure repetitive work itself can also be executed without database access. Product Owner argues, "In my opinion, we should not probably give database access to any external personnel. So, that might pose some limitations, and build dependency on the people that have the database access and people that would be doing the repetitive tasks." As a solution, CISO proposes, "In case of outsourcing, the simplest way of organizing the tooling would be developing one separate tool designed for the outsourced data transformation work. So that, e.g., classification would become as own process separated from the other software. It could require quite a bit of development work, but I do not see many downsides in that model. It would decrease our information security requirements towards the vendor." Having purely separate tooling for the third-party work would not only be beneficial information security-wise, but also it could decrease the pressure to cope with all the new emerging software features. CISO also argues that the clients would favor this approach, "It would be more acceptable to give an access to a separate classification tool instead of allowing wider rights, that are difficult to monitor."

Usually, technical development ideas of the data transformation tools emerge from the end-users, meaning DM specialists. If a third party would use Sievo's tools, there should be a way of noticing and gathering improvement ideas also within the new operating model. Continuous service DM Lead says, "There is also a need for feedback loops between Sievo's classification and consolidation champions and third-party employees in terms of product development requirements." Similarly, Head of People & IT is aligned with that, "At least from my point of view, it's kind of crucial that we do a part of the repetitive work in-house anyways, because we also need to develop the tools with which it's done, and maybe even, or probably even also continue automating some of those tasks. If an outside entity totally does the tasks, then it's really hard to develop them." On the other hand, the Product Manager underlines that it will depend a lot on the type of partnership. She says, "If we're lucky enough to get a nice partner that is smart enough to provide us also relevant feedback, then is just another source." The Product Owner is aligned with the Product Manager, "I see nothing stopping us testing something with the current tooling. So, I see more about the challenge of coordinating the work and requests."

It is also essential to notice that the limited DM resources create significant pressure on the product development to come up with new functionalities and better algorithms. The product Manager suggests, "If the supplier can do the same task in the same time frame for much fewer costs. So, we don't need to automate anything or improve any software. Maybe we don't need to do anything else on AI, learning settings improvements; it's just something that happens. And people manually check every two days the settings, and we don't need to improve it any further. And that frees up three engineers."

4.2.2.4 *Impact on Customer Perception*

As seen before, certain individual tasks have an impact on the customer perception, e.g., classification speed and quality are the crucial factors that the client is looking at the PoCs. However, the impact of outsourcing can also be analyzed from other angles than the perspective of different tasks. Usually, business people from the client side are not intensively interested in the ETL process; they rather just care about the correctness of the numbers on the dashboards. According to the Head of Sales, "The main thing is that the data is classified according to SLAs, it's correct, and the classification coverage is high enough."

Procurement analytics is characterized by handling massive amounts of clients' sensitive financial data in Sievo's databases. In the case of outsourcing, it most likely raises more questions than current Finland and the US-based in-house data processing. Head of Sales argues that "Outsourcing would cause difficulties in the sales negotiations. Clients are not going to take outsourcing well. The question rather is, how negatively they are going to take it and what type of additional negotiation work it will create." When it comes to customer perception and location, the Head of Sales also adds, "Most likely Europe would be completely fine, but Asia could be troublesome from the sales perspective. On the other hand, many of our competitors already have a team in Asia."

Since customer service is one of Sievo's core competencies, in the case of outsourcing, the company should guarantee that it is shown to the client as one Sievo. CEO of Sievo Inc. says that "It can't be in such a way that those external people that do that work are not seen as Sievo people. So that client would see a difference between the work coming from the third party and Sievo." Sievo must ensure that the third-party DM Specialists will care about the clients in the same way as the internal employees do. Head of People & IT says, "It is possible that outsourcing decreases the ownership of the work.

Ownership means sort of not just following the metrics, but actually trying to, for example, make the customer happy, or really caring about the work."

At this stage, we are still considering only internal classification validation to be outsourced, not external classification validation. If one day Sievo also decides to outsource the external validation, it would be even harder to guarantee that the clients would see the third-party DM Specialists as Sievo as well. In case of outsourcing reclassification and reconsolidation work, the external party should have some sort of access to operate in Sievo's ticket system. Additionally, a governance model should be developed for managing and guaranteeing frictionless communication between the three different parties: Sievo, clients of Sievo, and third-party DM Specialists.

4.2.2.5   *Relative capability position & Contribution to competitive advantage*

As the main principle, organizations should not outsource processes or tasks crucial for the competitive advantage and unique capability to perform it. However, outsourcing is a feasible option if the outsourcing partner has a superior ability to achieve a competitive advantage than the organization itself. (Mclvor 2008.) CEO of Sievo inc. argues that "I wouldn't transfer the data profiling to be very honest. So, I think that's a core competency. And by that data profiling, I mean this implementation heuristics set-up, so classifying the high spend items and getting that in place. I think that's something that we do extremely well, and we have a competitive advantage." It means that Sievo has a competitive advantage on implementation spend classification work. The company is doing it in a superior manner, which indicates that the task necessarily should not be outsourced. Excellent performance in a task that is highly contributing to competitive advantage signifies that the organization is capable of differentiating the product or service in the eyes of the customer (Mclvor 2008). Other than implementation classification work, repetitive data transformation tasks do not demonstrate a high contribution to competitive advantage coupled with strong relative capability position; hence all the rest of the tasks are transferable to an external party from that perspective. Indeed, external partnering could even facilitate learning new best practices and increase the relative capabilities of internal classification validation and consolidation validation work that is not currently conducted in a superior manner. Product Owner mentions, "I think I would be curious at least to know if they have any AI capabilities."

If the outsourcing would facilitate having DM resources at a lower cost, it could ameliorate Sievo's position compared to the competitors. At the moment, Sievo is a top-

end BIaaS solution with above-average pricing. Head of Sales mentions, "We have competitors, who are offering their service without additional implementation fee. So, suppose our DM resources were more cost efficient. In that case, we could either decrease our pricing becoming more competitive, or increase our profitability by having same income at the lower production costs." On the other hand, lower production costs could facilitate increasing the total DM resources. Those could be directed to increasing Sievo's competitive advantage. Head of Sales says, "The duration of the implementation phase is one of the parameters that clients use for evaluating different solution providers. It would be great if we could decrease the total duration of implementation projects." Moreover, if the outsourcing allowed the increase of total DM resources, some could be allocated to increase the current service level. Continuous service DM Lead mentions, "If we had more people, it would allow us to process the data even faster, meaning that the monthly process instead of three days would only take one day. So, these kinds of factors are also significant for the competitive advantages, because at the moment we need two to three days, depending on the amount of data sources. The customers are happy with that, but most customers want to have data out faster and faster. And if one day we want to offer weekly process for most customers, if not a daily process, we need some scalability." Furthermore, additional DM resources could also facilitate decreasing the reclassification and reconsolidation execution time. Customer DM Lead says, "We could leverage external service provider for diminishing the ticket response time, e.g., from current five days to three days." Overall better service level on a continuous basis could contribute positively to Sievo's competitive advantage.

4.2.2.6   Potential for opportunism

When outsourcing the operations to an external party, there is room for opportunism, and the relative strength between the buyer and supplier might distribute unevenly over time (Mclvor 2008; Mohr et al. 2011.) Therefore, it is crucial to evaluate the data transformation tasks' potential for opportunism. In the big picture, data transformation work is part of Sievo's core offering, and therefore, there is a general overdependency risk related to outsourcing the tasks. Continuous service DM Lead comments, "If for whatever reason the supplier is having resource challenges, we are pretty much dependent on those." The risk of becoming overdependent on the service provider increases even more in the long-term period if the company completely loses the capability of doing the data transformation work by itself. The continuous service DM Lead describes, "If the contract would

end, e.g., after three to five years, and the data management team as it is today doesn't exist anymore on Sievo side. Boy, we need to hire, and we need to hire in a short period of time, there's no cover-up this we would need to do a tremendous amount of communication with the customers on different levels and do a lot of prioritization which one goes first. And that would require that a lot of senior members that have had some point classification exposure and that means other projects inside the company might be compromised for that." It means that in the case of outsourcing, the company should have some sort of plan for protecting itself from opportunism and losing the skills of doing the repetitive data transformation tasks completely.

When evaluating the potential for opportunism on a more detailed level, the interviews showed that only classification has a high potential of creating overdependency on the supplier. The other tasks are not prone to that. Implementation DM Lead explains, "There is a risk of becoming overdependent on a single supplier when it comes to the classification work. Classification is our core competence impacting all the other modules of the software. If we lost the skill of doing classification, the supplier would have power over us." She also underlines the fact from the product development perspective, "We are not able to develop our tools if we do not know how they work. For product development, it's very important that DM Specialists actually tells them how they use the tools."

From a technological perspective, the risk is related to the tooling. If the external service provider uses the organization's tools, then Sievo is in control, but if the Sievo becomes dependent on the supplier's tools, there is a risk of creating overdependency. The Product Owner states, "If the external party is using our tools, then you can always take back the ownership and start using them ourselves. But if they are providing the classifications, consolidation, and some other means and we remove all our capabilities of doing our own classifications. Then we will be very dependent on them." To minimize the risk of vendor lock-in, it would be better to give access to Sievo's tools rather than becoming dependent on the supplier's tools. In the worst case, the Product Manager says, "I would presume we would only kill all of our tools if we pushed all of the efforts to an external party."

4.2.2.7 Required expertise

When it comes to required expertise, in the best-case scenario, outsourcing would not require heavy investments in knowledge transfer. DM Specialists says, "In the optimal situation, we have documentation so well presented that anyone who looks at the

documentation and has an understanding of the tools can work efficiently. It means that the data transformation tasks do not require years of previous experience, but knowing the tools and client-specific information is the most important expertise. However, the documentation is not yet at the level that all the classification tasks could be executed based on the written materials. Still, information sharing between different internal stakeholders is required. It will also be a pain point, especially on the continuous service side in the case of outsourcing. Classification Champion says, "We have some very demanding clients. I can never see it [outsourcing]. Even now, the knowledge transfer among us is troublesome. How confident are we that we can do it with a stakeholder that never worked with us before?"

On the contrary, there is historical evidence that required client-specific expertise is transferable in the implementation function even though the DM Specialists did not participate in all the client meetings. Implementation DM Team Lead states, "Knowledge transfer is not impossible. In the past, DM Specialists did not participate customer meetings; it is possible to transfer essential client-specific knowledge with one training instead of being continuously present in the client meetings." It means that at least in the implementation organization, the required knowledge and expertise would be quite easily transferable to an external party. In contrast, the continuous service organization would require developing a better information-sharing system for client-specific knowledge.

In addition to customer-specific knowledge, outsourcing would also require comprehensive knowledge transfer from the tooling perspective. If the external employees are using Sievo's data transformation tools, it will require extensive knowledge transfer. It is not only about knowing how to input a classification or consolidation decision to the system but also understanding how the whole system and AI behind the decisions work. Spend Product Manager says, "So, whoever would be taking care of these repetitive tasks would also need to understand the complexities of how the AI algorithms that we have work and what the settings should be to provide accurate decisions. It is not just about getting up to 95% coverage, but that there should be some quality check and some in-depth check of those decisions, right. So not just the human but the AI. We would need kind of long-lasting relationships with the supplier on two angles, the customer understanding customer-specific understanding, but also on the tooling, which is fairly complex on its own as of today." On the other hand, the exact requirement does not apply to the consolidation validation. Product owner says, "I see that consolidation use case is an

easier one, that fixing consolidation mistakes with our tool would be a simpler thing to crack. Overall consolidation is a lot simpler to outsource."

4.2.2.8 Required management capability

At least at the beginning of the outsourcing journey, the complexity of governing the DM work would increase. There would be some adaption period when both parties, Sievo and external service provider, would become acquainted with each other. During that time, DM Leader would play a key role in ensuring the frictionless transition. Continuous service DM Lead estimates, "If it's like a completely external service provider, we need to adapt because they also have standard processes. So, I am not entirely sure how much that will leverage our time, but I know for sure that it will require a lot of adaption. I see this will be a significant day-to-day work, pretty much like a professional service implementation, where we discuss with the customer on a daily basis about status updates. And that can range anything from three to six months, and then it might change to weekly managerial cooperation." According to continuous service DM Specialist, "We would now have one more stakeholder, like, in the whole chain of information sharing." It would impact not only continuous service but also implementation function as well. The implementation function is a matrix organization, which means that DM specialists' performance directly impacts the ongoing implementation projects. Implementation DM Team Lead underlines, "In general, management effort would increase due to project-based working. The project schedules are very tight, so we should reserve some security buffers if we need to fix some errors [cause by a supplier]." It means that managerial effort is needed to surveilling the quality of the work and make sure that the project schedules are respected. As seen before, client-specific information can be a crucial prerequisite for executing the spend classification successfully; therefore, DM Specialist states, "If we want to use another party supporting us in classification work, we need to make sure that we are not losing any information in knowledge transfer." It would require additional management efforts to optimize the crucial information sharing when having one new stakeholder as a part of the operating model. It is aligned with the continuous service DM Lead's statement, "There will be managerial tasks, such as setting up meetings with the external service provider on a frequent basis between the data managers and the people who have done the classification and the consolidation."

On the other hand, in the long-term, outsourcing could allow decreasing the required management capability. Implementation DM Lead says, "There should be some

sort of implementation period, which would require additional managerial effort, but after that, the situation could stabilize. Partnering could offer stability to the implementation project management since there would not be other tasks taking the time of the DMs, and we could trust the task execution in agreed time frame avoiding negative domino effect." Continuous service DM Lead says, "This would also help us in the long-term to kind of like forget about like training needs." In the case of outsourcing, especially internal classification and consolidation validation work would not necessarily require additional managerial attention than in the current operating model. Implementation DM Lead concludes, "Those are less complex tasks. Internal classification and consolidation validation can be executed without direct client contact; the tasks are not that client dependent, which would mean that the total managerial effort would remain the same."

4.2.2.9   Required HR Capability

Outsourcing in the long-term can facilitate decreasing required HR capability. Recruitment responsibility would be transferred to a third party. Head HR says, "I think it's up to the entity from which we're buying the services to find the workforce. So, if they have that, then it's not sort of our problem." Additionally, the third party would be responsible for the salary and compensation of those employees. Head of People & IT clarifies, "It's a different company, and they have their own offering. They probably have maybe a little different value proposition that people have chosen."

   Even though Sievo would no longer be responsible for hiring third-party DM Specialists, those would still affect everyday working life. Therefore, the Head of People & IT underlines that also in case of outsourcing, the company must understand the cultural factors when choosing the vendor. In the end, those people would represent Sievo for the clients and would be responsible for providing high-level customer service.

4.2.2.10 Information Security

Similarly, to Dhar & Balakrishnan (2006) and Oshri et al. (2009), in several interviews, people from different parts of the organization raised their concerns regarding clients' information security in case of outsourcing. The company's CISO clarified which information security factors should be considered in case of outsourcing the data transformation activities. There are multiple interfaces and data points that DM Specialists use in their work on a daily basis. In outsourcing, technological capabilities are not necessarily

restricting, for which tools Sievo would give access for a third party. CISO states, "Technology does not necessarily limit us, but rather the workload and the promises given to the customer regarding the data processing. It [outsourcing] is rather business decision than information security decision." The information security level should remain the same even though a third party conducted the repetitive data transformation tasks. CISO argues, "Of course, the starting point is that we choose a vendor that is trustworthy, has well-functioning processes and views information security with the equivalent seriousness than we do."

Currently, DM Specialists have access to the databases, not only the graphical classification tool, since they often have to conduct in the data non-repetitive more complex tasks. However, CISO raises questions, "What data point should the person conducting data transformation tasks have in use? Can those data points be somehow restricted?" According to the data minimization principle, the third-party employees should have access only to the tools and data points that are necessary for conducting the data transformation tasks; consolidation validation is possible to do by using vendor name and address, the actual spend is not a mandatory data point. CISO underlines, "If we gave database access to a third-party, it would be difficult to limit the datapoint that they could access. Additionally, saving the data for their own purposes would become easier." It means that giving access to the database would increase the data loss risk, and it would be challenging to follow the data minimization principle. Moreover, database access would require allowing a third party to enter the organization's internal network. CISO comments it as following, "It would be a quite big step if we gave access to our internal network. Not giving access would be the most clinical approach. In case we want to give it, we should start rethinking the network topology, which has been quite simple so far." On the other hand, it is possible that in the future outsourced DM Specialist do not conduct only repetitive work but also all types of DM work. In that case, they would sometimes need to, e.g., send data from the database to the end-users, which could maybe justify the database access. In that case, CISO argues that "We should almost start evaluating third-party employees in the individual level for guaranteeing their suitability for handling clients' sensitive data."

When an increasing number of employees access clients' sensitive financial data, it exposes the company to greater insider risk. Especially in outsourcing, CISO sees more significant risk, "Outsourcing brings a new aspect when the employees are not working for Sievo anymore. We cannot give the same trust to third-party as to our own employees

since there is no same type of loyalty and fidelity towards the employer. We must have a model for decreasing the insider risk before starting the partnership in practice." CISO also adds, "In case of outsourcing, we should do some sort of investments on the data loss prevention (DLP) products, which monitor how the data is processed and printed out from the system. Those tools also prevent malpractice. We should definitely have one in place, and this type of investment should be prioritized." The other cost would increase from updating the existing contracts. Head of Sales mentions, "our contract appendix 5 lists our sub-contractors and third-parties. The company must update its existing contracts if it starts buying data transformation work from an external service provider."

One of the most critical questions would relate to data hosting and accessing locations in case of outsourcing. CISO says, "The location is a hot topic also for our European clients due to tightened GDPR. We have promised to quite many clients that their data is kept in the European Economic Area (EEA)." It is also aligned with the Head of Sales' comment, "My biggest fear is, how the clients will react to his and information security matters. I already know that they will not take it well, but the question rather is, how bad reactions they will have. It will require some additional work." In case of moving data transformation activities outside Europe, it will require negotiations and updating current contracts with the existing clients.

Outsourcing would also limit the visibility of the current state of information security. CISO mentions, "We have promised to our clients that our computers' antivirus protection is up to date. Currently, we can easily monitor it in a centralized manner. Still, in the case of outsourcing, we would lack visibility completely. It should be managed some other way." IT auditing could serve as a mean guaranteeing the supplier's sufficient level of information security. CISO argues, "We are about to outsource such significant tasks that we should choose a vendor that allows us to audit their processes. We would make sure that they are operating as they promise to operate." Additionally, the external service provider should have the same information security certificates as Sievo has. In the end, CISO concludes, "If those [IT auditing results and certificates] are in order, I do not see that outsourcing would significantly worsen our information security."

4.2.3   Changing the production location: nearshore and offshore in-house production

Instead of outsourcing, the organization is also considering establishing a new entity dedicated to DM work and especially data transformation tasks for gaining some cost reductions. From a qualitative perspective, the insourcing scenario is closer to the current state.

In Table 8, different data transformation tasks are evaluated against the criteria similarly to the current state and outsourcing. Color coding in the evaluation table demonstrates if the value is more desirable (green), the same (no color), or less desirable (red) compared to the current state.

4.2.3.1 Efficiency

The most relevant factors impacting the efficiency of the repetitive work, such as poor incoming data quality and client specificity, do not necessarily change when changing the location. If the work was organized differently, allowing specialization in the new entity, it could offer some efficiency improvements. Implementation DM Lead mentions, "If we had an employee, who was very good at validating, e.g., marketing category and does it very often, the working becomes more efficient." When it comes to consolidation validation efficiency, if an employee did mainly only that task on a daily basis, it could have a negative impact on efficiency. Implementation DM Lead says, "It is impossible to do consolidation validation work very efficiently many hours in a row." Additional planning efforts should be made to guarantee the sensibility of the work despite its repetitive nature. Insourcing could also allow restructuring the reconsolidation work in the implementation phase, meaning that the consolidation would not be done on a project-basis. Implementation DM Lead points out, "Supplier consolidation could be organized in the more efficient matter - not anymore project-based, which would allow avoiding the double consolidation validation work."

From the more general operational efficiency, at the beginning of establishing the new entity, a short-term efficiency decline is unavoidable. Head of People and IT says, "If there is a separate entity handling a part of the work, then it requires coordination, and that typically reduces efficiency." In order to cope with the complexity, establishing the new entity should be done in a coordinated manner. Continuous service DM Lead states, "I think from the operating efficiency will drop down a little bit at the beginning of the transition period, that is for sure. So, we need to kind of find a way to continuously ramp this up. We hire five to ten people, and then we start with them and not like one or two people, and we ramp it up slowly. That is the only thing I see."

Table 8 Insourcing the repetitive data transformation task

Task	Effi-ciency	Quality	Tech. de-velopment capability	Pos. Im-pact on Customer Perception	Pos. Contri-bution to Competitive Advantage	Relative Capabil-ity Posi-tion	Required Expertise	Required Manage-ment Ca-pability	Re-quired HR ca-pability	Required Inf. Sec. Capabil-ity
1. Implementation: Spend Classification	3	3	4	4	4	4	2	3	4	3
2: Implementation: Classification Validation	3	3	4	3	2	1	1	3	4	3
3. Implementation: Repetitive Reclassification	3	3	4	4	4	3	1	3	4	3
4. Implementation: Supplier Consolidation Validation	2	2	4	2	2	1	1	3	4	3
5. Continuous Service: Spend Classifications	3	3	4	4	4	3	2	3	4	3
6. Continuous Service: Repetitive Reclassifi-cation	2	3	4	4	3	3	1	3	4	3
7. Continuous Service: Consolidation Valida-tion	2	2	4	2	2	1	1	3	4	3
8. Continuous Service: Repetitive Reconsoli-dation	2	3	4	4	3	2	1	3	4	3

4.2.3.2 Quality

Correspondingly to operational efficiency, insourcing would not indeed have a significant impact on operational quality. Compared to the current state, most likely, some additional effort should be allocated to quality management. Implementation DM Lead says, "At the moment, we do not have clear classification quality metrics. Insourcing could require building more clear guidelines." The other DM Lead agrees, "Insourcing impacts on how we validate the data, how we make checkpoints." On the other hand, insourcing would allow better tools over the data transformation work quality than outsourcing.

If the insourcing allowed reaching cost-efficiencies, the company could allocate more resources to the data transformation work, which would increase the decision-making quality related to classification, classification validation, and consolidation validation. It is the same thing that applies correspondingly to outsourcing, but insourcing would allow the usage of all the best practices that Sievo has developed over the years. The organization would not be dependent on the external party.

4.2.3.3 Technological development capability

In the interviews Product Owner and Product Manager did not have strong opinions of insourcing, which indicates that the technological development would continue in a similar fashion than currently. The location of the data transformation tasks does not play a significant role from the technological perspective. The Product Owner says, "The location does not necessarily matter from the technological point of view if the employees have a good internet connection. They can be located wherever there is a network connection. Then the other thing to consider is the time difference, that how much collaboration it requires with other Sievo people."

Having Sievo owned new entity would facilitate that the fruitful cooperation between the DM teams and product development continue in a similar fashion since it is a common interest to develop the existing tools and their features. Even though insourcing would not eliminate the automation needs completely, the additional DM resources could decrease the high pressure directed towards the product development. In that sense, the insourcing model would be the optimal choice from the product development capability. There would not be a risk of killing the current tools and becoming dependent on the

external data transformation means. Insourcing would enable internal learning, but at the same time, it would lower the pressure of automating the repetitive work.

4.2.3.4 Impact on customer perception

From the sales perspective establishing a new entity is a more attractive option than out-sourcing. Head of Sales says, "Total outsourcing is more challenging to carry out, so from the sales perspective, the best option is setting up a new subsidiary in Europe. On the other hand, the other options are not impossible." Also, DM Team Leads are aligned with the Head of Sales. Implementation DM Lead concludes, "If the repetitive data enrichment work was executed at other Sievo entity, I do not believe that it would impact on our current customer perception." According to continuous service DM Lead describes, "it is foreseeable." If other criteria such as quality increase through insourcing, it most likely also improves the positive customer perception.

Also, insourcing can be expected to guarantee the same top-end customer service that the current DM Specialist are offering to the clients. Suppose the DM employees are part of the Sievo. In that case, they can less frictionlessly cooperate with the clients and internal stakeholders such as implementation project teams or Customer Success Managers, which most likely facilitates better end-user service. In the end, that customer service is one of the company's core competencies. CEO of Sievo Inc. describes, "Clients are not expecting zero-defect service. They are expecting a person who they think care about them, and who keep them up-to-date and who they can trust that takes care of them no matter what." A strong customer service mindset is more easily transferable to internal employees guaranteeing outstanding customer perception despite the insourcing.

4.2.3.5 Contribution to competitive advantage & relative capability position

Insourcing operating model would allow transferring all the unique ways of working to the subsidiary. It would guarantee that Sievo does not lose any existing superior capabilities related primarily to spend classification. Correspondingly, Sievo is not able to access any new data transformation capabilities, which could be beneficial for tasks that do not demonstrate superior performance compared to the competitors, such as consolidation validation. In the case of insourcing, Sievo mainly continues operating with its currently existing capabilities, which as a principle is a positive thing. In the end, Sievo's high growth demonstrates that the current model is successful.

Insourcing could increase the competitive advantage if the cost of data transformation activities decreased by changing the location to a more economical country. The same type of positive impacts on the competitive advantage can be expected from insourcing than outsourcing, given that the total amount of available DM resources increases. Implementation DM Lead describes, "Having more resources with the same costs would facilitate increasing the output quality, and operational efficiency." Continuous service DM Lead agrees, "Insourcing would allow that we can provide an even better service than we do today to the customers. For example, depending on the location, but we could even promise that the classifications or the feedback implementation can be provided in the client's respective time zone." The assumption is that the insourcing would allow allocating more resources to the most critical competitive advantages strengthening Sievo's position in the procurement analytics BIaaS market.

4.2.3.6   Required expertise

When it comes to the employee's required expertise, the exact basic requirements will apply than in the current model. Also, if the candidates were hired mainly for the data transformation work instead of cherishing the idea that DM Specialist is only an entry-level position, and the employees are rapidly moving forward in their career with Sievo. Implementation DM Lead points out, "If instead of moving quickly forward with their careers, employees would stay longer time in DM role. It would allow deepening the expertise and leading to better performance." On the other hand, the monotonic nature of repetitive work most likely impacts employee satisfaction, decreasing the average length of employment. It instead has a negative impact on the likelihood of building up long-term role-specific expertise and efficiency improvements. Moreover, changing the career development objectives would allow modifying the hiring requirements. Implementation DM Lead mentions, "Insourcing would enable us accessing new beneficial language skills and industry knowledge."

Concerning client-specific knowledge requirements, the expectation is that the insourcing model would facilitate more frictionless information sharing than outsourcing. As seen before, especially in the continuous service function, there is still undocumented silent information, which needs to be shared to perform the classification tasks successfully. Furthermore, if only part of the repetitive data transformation tasks were transferred, it would increase the complexity of required expertise transmitting. Implementation DM Lead says, "Classification validation is heavily related to actual classification,

and the same person usually executes both. If they were separated, it would not be impossible, but it would require arranging an efficient communication channel between the people working with the classification and classification validation."

4.2.3.7 Required management capability

In the case of setting up a new entity, it would increase the required management capability. According to continuous service DM Team Lead, "At the beginning of the project management tasks would be in increase. My assumption is something from three, six, up to 12 months, because I see the knowledge transfer and setting up processes will require a good proportion from the management aspect, but also the existing DM Specialist." The existing DM Leads' effort, as well as some of the DM specialists' resources, should be allocated to the insourcing project. It is aligned with the Head of People & IT, "I guess there would be a lot of new stuff like managing, let's say, resourcing, or the interface between the current teams, for example, data management of the implementation function and the new entity."

Even though the required efforts would increase in the short-term, the model would not differ from the current situation significantly. The company already has existing processes and ways of working in place; those would be just transferred into a different location. DM Team Lead describes it, "If we set up the whole operation center in-house, it's a more streamlined process. We know how to train the people. We know how to associate them with our company. They're in the company hierarchy so that the whole task will be manageable." The entity could be even managed more independently than the other existing entities. Head of People & IT describes, "' Make' option in another country; it would be running, as a quite independent entity, more independent than, for example, the U.S currently. We are now managing two active entities. I don't see adding a third one there, that it would bring a lot of different sorts of new requirements." Of course, some minor changes to the current organizational structure should be most likely made. The Head of People & IT comments, "It would be easier if there was one point of contact or somehow like, if there was one counterpart at Sievo handling the communication and planning, etc., with the new entity. So there, I don't know if it would make sense to do some changes to the current organizational structure as well."

4.2.3.8 Required HR capability

In the case of setting up a new entity, it would require lots of additional HR, IT, and legal effort, especially at the beginning of the project. There would be lots of planning to do, and location would be one of the decisions to be made. The required efforts can be divided into two categories based on the timeframe: short and long-term investments. There is lots of additional work to do in the short-term, from choosing the new location to hiring people and training them. In the long-term, additional efforts are required to incorporate the employees part of the company culture, take care of the administrational factors, and take care of the employees' well-being.

When choosing the location for the new entity, Sievo is searching to obtain cost savings with the operating model changes. Head of People & IT comments it as following, "If we started a subsidiary, then I would make sense to locate that in a country where the salary level is, is lower." There are also legal questions to be analyzed related, for example, to competition clauses. Head of People & IT comments, "It would make a location or country more lucrative if there were, for example, non-competition clauses. If we hire people there, they wouldn't go straight away to some competitor and take the Sievo insights with them. Moreover, cultural factors must be taken into consideration. Head of People & IT comments, "If we would start a sort of subsidiary or, then buy services from countries that are culturally very far from Finland, it's probably going to be a learning curve to work with that entity." However, Sievo is a very multicultural company, and Head of People & IT mentions, "We have at Sievo already people from like 30 countries. So, if we have knowledge of a certain country and culture already in-house, it might be beneficial when establishing a company." Moreover, the political situation of the destination country should not be overlooked. Head of People & IT says, "If considering going to a different country, especially starting something Sievo-owned there, then looking into the political situation there would be a part of the decision-making process."

After analyzing the relevant factors and making the location decision, there would be a significant amount of administrational work expected from the people & IT function. Most likely, the finance function would establish the legal entity, but People & IT function would be responsible for arranging all the employees-related matters. Those kinds of services can be either bought from an external service provider or hire one HR person for the new entity. Head of People & IT describes, "Making sure that we meet all the requirements posed to employers in that country would probably mean we would need to find a

partner in the country. A company there that is taking care of many of the traditional HR tasks, e.g., payroll, but also like handling benefits, keeping the employee data and so forth."

In the long-term, one of the critical factors is employees' well-being. The monotonous, repetitive work has created frustration and made some employees think about quitting their job. Implementation DM Lead says, "In case of setting up a new entity, the same well-being questions would follow or even increase, especially if the people would not be offered balancing development work." On the other hand, in the case of insourcing, the hiring policy could be modified, corresponding more the specific tasks. Implementation DM Lead argues that the hiring policy should be adjusted to focus on specific DM roles than overall Sievo, "Employee does not need master's degree for executing the repetitive data transformation tasks." On the other hand, if the employees were hired for a narrow role, not offered the same career prospects and same value proposition as the other employees in U.S and Finland, that could negatively impact the company culture. Head of People & IT thinks, "If we then create an entity where the people are not sort of a part of that one Sievo anymore. I'm just making some assumptions now, but if the employees were not offered the same opportunities and benefits as the other employees, It is, to some extent, creating a sort of second class. I think that would have an impact on the culture. Should we not do the insourcing because of that, I would not necessarily say so. But we at least need to communicate that the value proposition would just not be the same for the people who work in that entity than for the rest of the company."

4.2.3.9 Required information security capability

It would not be the first time establishing a new active Sievo entity, and therefore adding one more entity would not be a significant shift compared to the current situation. The company is already serving US clients and conducting their data transformation tasks in the US: CISO describes it as following, "We operate in the same way in Finland and US. Data protection regulations are stricter in Europe, and therefore US DM Specialists take care of only US clients, whereas the rest of the clients are handled in Finland." In the case of setting up a new entity outside EEA, data hosting and processing locations are as critical factors as in the outsourcing scenario. Even though Sievo would not need to revise appendix 5 concerning third parties, the contracts should still be updated. CISO mentions, "Our contract package includes Data Processing Agreement, which relates to strong personal data protection, which defines where and how we can process clients' data."

Legally, there is no fine line, whether data can be accessed and processed according to GDPR if done outside EEA with remote access. This topic requires more investigation with legal experts.

Even though the company already is experienced having an active entity, it is still vital to evaluate critically if the employees executing the repetitive data transformation tasks need the same access rights as the current DM Specialists, who conduct a wider range of tasks. Indeed, CISO argues, "Even if we would have our own subsidiary, from my perspective the clearest approach would have one separate classification tool for those employees." According to Spend Product Management, employees' access rights could be quite easily limited for required tools, "Already now, you can create a user that has no access to any analytics and only has access to the classification or wherever he or she has been assigned to. I guess that that could be supported." Similarly to outsourcing, CISO encourages critically evaluating what tools and data points are necessary to have access to perform the data transformation tasks successfully. However, when it comes to database and internal network, for the internal employees, it would be easier to grant access to those environments, if in case that would be required at some point.

4.3   Which

According to TCT, when making the outsourcing decision, the main objective is to minimize the occurring costs and maximize the benefits. Naturally, some of the costs and benefits can be applied for all three scenarios, e.g., processing client's sensitive data outside the EEA area can be against clients' will and increase the resources required for sales negotiations. However, this chapter summarizes the most prevalent costs and benefits related to different scenarios for supporting the final decision-making. Quantitative costs and benefits refer to direct inducements of data transformation production and governance costs. Qualitative benefits and costs instead indicate non-quantifiable positive and negative consequences related to different scenarios. Table 9 provides the summary of the findings.

Table 9 Cost and benefits of the different scenarios

	Current State	Outsourcing	Insourcing
Quantitative costs	HIGH: - Current operations in high-cost countries: Finland & US	LOW: - Cost is contractual - New infosec investments: DLP, tool modifications, IT	MEDIUM: - Establishing an entity requires significant investments

(Optimal low)	- Continuous need of hiring new DM Specialists :	auditing, increasing risk management - Updating the existing contracts with the clients - Increased managerial effort in the short-term	- HR: partner in the destination country, establishing new policies - Small infosec investments - Increased managerial and administrative effort
Qualitative costs (Optimal low)	MEDIUM - The amount of DM resources is slacking compared to the growth, resulting in negative consequences - Not much production flexibility during the peak times - Repetitive tasks unmotivating to the DM employees	MEDIUM - Potential for opportunism - Risk of losing relative capability position in the tasks that are directly contributing to the competitive advantage - Negative impact on customer perception, esp. offshoring	LOW - More challenging to offer the same company culture for the employees, difficult to offer versatile career prospects - Offshore insourcing would require more negotiations with the clients - Employee well-being questions
Qualitative benefits (Optimal high)	HIGH: - All the processes optimized for the current operating model - No risk for opportunism - Data processing taking place in the mother company in Europe is a good selling point - Flexible organization capable of managing complex clients - Company culture, the feeling of family - DM employees contribute actively to product development	MEDIUM: - Lower costs could facilitate increasing the total DM resources - Possibility to scale the production during the peak times - Less managerial responsibility in the long-term - No responsibility for the knowledge transfer, career development, company culture - Possibility to prioritize the suitability of employee profile for the specific task, not the whole company culture - Access to new best practices	HIGH - Lower costs could facilitate increasing the total DM resources - Stronger control over the output - Facilitating close cooperation between DM and product development, less automation pressure - No risk of opportunism - More loyal employees - More acceptable option from the client perspective - Easier to transfer additional tasks in case the operating model changes turn out to be well functioning

4.3.1 Minimizing the costs

4.3.1.1 Quantitative

According to CFO's calculations, the most expensive option is the current state because of the high-cost locations and overall high salary level compared to tasks' difficultness. If the simplest tasks are transferred to the lower costs country, it decreases the data transformation production costs.

Establishing a new entity is a cheaper option than continue with the current model, but it still requires investments. As stated by CFO, the insourcing model would require notable one-time investments related to legally setting up the entity, having an HR partner in a new location, hiring and training the new employees. Most likely, some small investments in information security would be expected as well. Some additional long-term costs would increase from supplementary managerial layer for coordinate and administrate one new entity and its employees.

From a financial perspective, outsourcing is the cheapest option. According to CFO, the production costs depend on the contractual price agreed with the suppliers. Usually, in addition to production costs, the organization must pay some sort of additional fee for the vendor. On the other hand, those expenses cover all the costs related to hiring and training the employees as well as executing the given tasks. In addition to vendor costs, outsourcing would require investment in information security and updating the contracts with the existing clients. There should also be efforts invested in creating a governance model for organizing the communication between the clients, supplier, and Sievo.

4.3.1.2 Qualitative

From the qualitative perspective, the highest costs of the current model are the consequences emerging from the limited resources. It is primarily shown as a trade-off in consolidation validation quality and tight implementation schedules. Limited DM resources do not facilitate high production flexibility during peak times, forcing the organization to sometimes using more expensive resources from the other functions. Also, in continuous service, the DM team has needed to delay different process development projects and prioritize the repetitive monthly and ticket work. In the current model, the product development is facing significant pressure to improve the automation level.

In the case of insourcing, the required HR effort would increase. If the employer's value proposition in a new entity is significantly different than currently, there is a risk of creating "second class" employees, which is against Sievo's current philosophy. It could harm the company culture and decrease the sense of family. Moreover, the challenges related to the monotonic nature of the repetitive work would also follow to the subsidiary, and Sievo should find a solution to that. In the case of offshore insourcing, the organization should guarantee GDPR compliance and conduct some negotiations with existing clients because of changing the data processing location.

Outsourcing would have an even more negative impact on the customer perception, especially in the case of moving the activities outside EEA. Most likely, the clients would not feel comfortable giving access to their sensitive business data to another external partner. Moreover, outsourcing would significantly increase the information risks, especially if giving access to Sievo's database. The organization would be prone to become overdependent on the supplier in several fashions. First, a third party would have the absolute control of the data transformation output, which could decrease the organizational capability to offer a unique customer experience. Furthermore, if the supplier used its own data transformation tools, in the long-term, Sievo could decrease its unique skill of doing classification and, in the extreme case, kill its own tools. Most likely external DM Specialists would have way more limited job descriptions than internal employees, e.g., most likely, the external DM employees would not participate that closely in developing Sievo's tool.

4.3.2 Maximizing the benefits

4.3.2.1 Current state

Naturally, from the qualitative perspective, the current model provides a lot of benefits. The high growth shows that the organization has managed to compose the winning formula in the procurement analytics BIaaS market. All the existing processes are optimized for the current operating model. Sievo has full control of the data transformation work output, and there is no risk of becoming overdependent on any external party.

As treating client's sensitive business data BIaaS, the current locations are often an essential factor for the clients in the sales negotiations. Moreover, the organization has managed to build well-functioning talent pipelines and information security capabilities

in the current locations. Having a simple organizational structure has also enabled creating one shared company culture and sense of family.

Even though the organization has had challenges in available resources during peak times, the company still has organizational flexibility to provide tailored solutions for demanding clients. Having everything produced in-house facilitates fruitful inter-organizational cooperation between different functions. DM Specialists and product development closely collaborate for creating new functionalities and improving the existing data transformation tools.

4.3.2.2   Outsourcing

Outsourcing facilitates significantly lower costs that instead would allow increasing the total DM results. When allocating more DM resources to the data transformation work, it positively correlates to the output quality. Furthermore, in the long-term, it would enable Sievo to increase its service level by decreasing lead and ticket solving time and, e.g., shifting the continuous data processing from a monthly basis to a weekly or even a daily basis. An external service provider could possibly allow increasing data transformation production levels during peak times. Overall, these improvements would positively contribute to the company's competitive advantage.

Outsourcing would allow also decreasing the overall training and HR efforts. The organization would no longer have responsibility for the employees' training, career development, and company culture. Also, the requirements towards the employees could be lowered when the external staff would be purely executing the repetitive tasks. When the supplier was in charge of the resourcing, it would increase operational efficiency from that perspective. Moreover, the external partnering could be a fruitful opportunity to learn new best practices and enhance the data transformation capabilities, which had no high relative capability position yet.

4.3.2.3   Insourcing

Insourcing is expected to increase the overall quality level if the cost efficiency would allow allocating additional resources to data transformation work. However, Sievo could still use its best practices and over the years learned lessons. Moreover, in the case of insourcing, Sievo would have more substantial control over the data transformation output quality. Sievo could expect the same customer service mentality from its employees

in the subsidiary than in the mother company. Increasing data transformation quality coupled with already existing strong customer orientation would most likely enhance the positive customer perception.

From the product development perspective, insourcing would guarantee close cooperation between DM teams and product development also in the future. Moreover, there would not be a risk of killing the own data transformation tools and becoming overdependent on a supplier. From the sales perspective, keeping the data in Sievo's possession is a significantly more acceptable option than outsourcing.

In the case of insourcing only the pure repetitive work, not the more demanding DM tasks, Sievo could decrease the hiring requirements, e.g., the data transformation employees would not need a master's degree to conduct the repetitive tasks successfully. Moreover, if the insourcing turns out to be a lucrative operating model, it is relatively easy to transfer additional tasks to the already existing subsidiary.

4.3.3 Recommendations

When evaluating the optimal operating mode, the essential factor is to consider that the current operating model and practices have already proven successful, generating over 30% annual growth over five years. The high growth just has showcased the limited scalability of the data transformation tasks, which has not been that evident with the smaller number of total clients. There is room for some technological automation efforts; however, the organization's strategical choice to offer customized solutions limits automatization. If the BIaaS company continues with the current operating model, the production expenses continue increasing, limiting the operating profit.

From the pure quantitative cost perspective, outsourcing the repetitive data transformation tasks would be the cheapest option. Also, the cost-efficiency would guarantee to increase the total amount of DM resources providing qualitative benefits. On the other hand, the outsourcing option includes more uncertainties from multiple perspectives, e.g., information security, customer perception, employee commitment, and opportunism. Moreover, Sievo would risk losing its relative capability in spend classification, organizational flexibility, and world-class customer service orientation, which have been an integral part of Sievo's success story.

Insourcing would allow Sievo mainly continue operating according to its best practices and superior relative capabilities. However, the company should be careful that it is not changing the location of the resource-intensive DM tasks without reaching cost

savings. A notable cost efficiency improvement is the prerequisite for increasing the total DM resources. It will be challenging because the current contracts oblige the data processing taking place in Europe. According to the company's CEO, the labor arbitrage is lower than in offshore locations. If the organization manages to keep the quantitative insourcing costs in control, that operating model has in total the most benefits and the most negligible costs. Therefore, results suggest that the BIaaS company should insource its repetitive data transformation tasks.

5   CONCLUSION

5.1   Fulfilling the research objectives

This action research essentially investigated the research question, "What is the optimal way of organizing manual, repetitive data transformation tasks in a Business Intelligence as a Service organization?" For arriving at an answer, the research process followed the stage model of IS Outsourcing by Dibbern et al. (2004).

First, the study started with analyzing the reasoning behind the potential need to outsource or insource the repetitive data transformation activities. Most of the negative consequences related to the current situation are emerging from the lack of resources. Ultimately, even though the consequences of the lack of resources are the motivation for the company and very visible in the organization's everyday life, they are not the actual root cause for the current situation. The primary root cause is the lack of scalability of the data transformation tasks. Currently, poor incoming data quality and client-specific customization restrict the automation, and extensive manual, repetitive work is required. The manual data transformation work cannot be scaled. Still, the workload increases with every new client leading the company to continuously hiring highly educated people for conducting the repetitive work in their current locations in Finland and the U.S. Potential benefits and risks related to the outsourcing and insourcing were translated to evaluation criteria for defining the optimal operating model option in the later stage of the project.

After understanding the root cause and establishing the evaluation criteria, the objective was to define the repetitive manual tasks at a detailed level. The procurement analytics and spend analysis-related data transformation tasks are spend classification, spend classification validation, and supplier consolidation validation. Additionally, the client feedback implementation, reclassification, and reconsolidation can be considered repetitive data transformation tasks as well. At the beginning of the project, the organization's top management defined four different scenarios to be compared against the current model to understand the best scenario. Those four scenarios were offshore outsourcing, nearshore outsourcing, offshore insourcing, and nearshore insourcing. However, the internal interviews showed that the comparison between the ownership of the tasks is a significantly more prevalent factor at this phase of the IS outsourcing project than the location dimension. Therefore, in the end, the evaluation was conducted between three different scenarios: the current state, outsourcing, and insourcing.

After selecting the different scenarios to be analyzed, defining the repetitive tasks, and choosing the optimal evaluation criteria, the objective was to obtain an in-depth understanding of how outsourcing and insourcing the data transformation tasks impact the BIaaS organization. Cooperation with an external service provider provides significant cost reductions facilitating an increase of total DM resources. Additional resources would increase general output quality, improve the service level, and remove hiring and training efforts. On the other hand, the outsourcing scenario would augment operational uncertainty, negatively impact customer perception, restrict internal technological development, and require more information security capabilities. Moreover, outsourcing the core competitive advantage with a relative capability position could harm the organization and expose it to the potential for opportunism.

In the end, the most optimal way of organizing manual, repetitive data transformation tasks in the BiaaS company is insourcing. The high-growth proves that the organization has already found the competitive advantages and superior ways of performing, especially in the spend classification. Insourcing enables enhancing the success factors by changing the operations to a more cost-effective country. The organization just must ensure that the new location indeed provides cost savings so that it is possible to increase the total amount of DM resources for realizing the expected benefits. The major practical contribution of this research is the detailed description of arriving at the optimal data transformation operating model recommendation that the case company can exploit in their final decision-making.

5.2   Limitations

Research validity is a classic evaluation criterion for understanding the strengths and the limitations of the research. The standard way of establishing the validity of the research is triangulation. Triangulation instead is a process of exploiting several perspectives for refining and clarifying the research results. Triangulation can be analyzed from several perspectives: methodologies, methods, data, theories, and researchers. (Eriksson & Kovalainen 2011.)

The research is from a methodological triangulation perspective limited. It is qualitative research, whereas mixed methods could have been utilized to increase the dissertation's validity. For example, the researcher could have conducted a survey or questionnaire among all the Sievo's (DM) employees to confirm the final findings of the qualitative benefits and costs within different scenarios.

Even though the research exploited several research methods, the data sources were only the company's internal ones. It means that all the interviewees were directly or indirectly linked to the activities to be potentially insourced or outsourced. At the beginning of the project, the employees were informed that nobody is losing their job due to potential insourcing or outsourcing; however, it is still possible that the results contain subjectivity. Moreover, some interviewees do not have previous experience in outsourcing or insourcing, which means that they have limited capabilities to envision the potential consequences of operating model changes. That is why having an external data source, e.g., benchmark interview, could have increased the validity of the research results.

When it comes to triangulation of theories, the research heavily relayed solely on the Dibbern et al. (2004) framework. At the beginning of the project, the researcher read lots of IS/IT outsourcing literature and chose the Dibbern et al. (2004) framework as the most suitable for this research. However, the decision was quite intuitive. At the beginning of the paper, there could have been a more comprehensive review of different IS/IT outsourcing frameworks for improving the validity of the research.

This paper had only one researcher meaning that the empirical materials were only analyzed by one person limiting the findings' validity. Anyhow, the project had two people continuously managing and monitoring it. Throughout the project, the researcher needed to justify the findings and provide the explanation for the other project manager, increasing the validity of the results.

5.3 Implications for the further research and academic contribution

In nature, the action research is iterative, and correspondingly this research could include multiple additional research loops. First, the study could be extended to investigate other operating models, e.g., a freelancing or mixed approach of insourcing and outsourcing, to understand the actual optimal operating model for data transformation work.

After choosing the best option, the process could continue to the recommended direction by following Dibber et al. (2004) IS outsourcing framework phase 2, which contains two sub-steps, "how " and "outcome." *How* step is related to vendor selection, relationship building, and relationship management. At the same time, *outcome* is about gathering the experiences and learning as well as understanding the types and the determinants of the insourcing success. Phase 2 could be carried out as an additional action research cycle. Furthermore, after completing the whole IS outsourcing or insourcing project, the realized benefits and risks could be compared to the results of this research thesis.

On the other hand, instead of investigating the most cost-effective option with minimum qualitative trade-offs, the organization could direct more resources to technological development. Text-based classification, more advanced classification validation, and supplier consolidation tools could also serve as a means of decreasing the manual data transformation work. In the end, automation is the most effective way of scaling the procurement analytics ETLA process-related work.

Even though action research as a methodology is not aiming to create universal knowledge, it is still important to acknowledge the implication beyond the understanding developed in the project (Coughlan & Coghlan 2002). The research serves as an example of conducting Dibbern et al. (2004) IS Outsourcing research phase within a BIaaS organization. The actual results are very case-specific, but they contribute to the procurement analytics field, which is still a narrowly researched area. The paper provides a detailed description of different procurement data transformation activities taking place during the ETLA process. Additionally, the research comprehensively portrays the potential emerging benefits and risks of outsourcing and insourcing data transformation tasks. Theory building in action research is incremental, meaning that the shift from situation specificity to universal knowledge happens in small steps. A similar study of conducting Dibbern et al. (2004) phase 1 could be carried out on the other BIaaS organization for understanding the benefits and risks of outsourcing and insourcing the data transformation tasks at the more general level.

REFERENCES

Abedjan, Z., Chu, X., Deng, D., Castro Fernandez, R., Iluas, I. F., Ouzzani, M., . . . Tang, N. (2016). Detecting data errors: where are we and what needs to be done? *Proceedings of the VLDB Endowment, 9*(12). doi:10.14778/2994509.2994518

Alaghehband, F. K., Rivard, S., Wu, S., & Goyette, S. (2011). An assessment of the use of Transaction Cost Theory in information technology outsourcing. *The Journal of Strategic Information Systems, 20*(2), 125-138. doi:10.1016/j.jsis.2011.04.003

Ang, S., & Straub, D. (2006). Costs, Transaction-Specific Investments and Vendor Dominance of the Marketplace: The Economics of IS Outsourcing. In R. Hirschheim, A. Heinzl, & J. Dibbern, *Information Systems Outsourcing* (2nd ed., pp. 27-56). Berlin : Springer, Berlin, Heidelberg. doi:10.1007/978-3-540-34877-1

Arunachalam, D., Kumar, N., & Kawalek, J. P. (2018). Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. *Transportation Research Part E: Logistics and Transportation Review, 114*, 416-436. doi:10.1016/j.tre.2017.04.001.

Asikainen, A.-L., & Mangiarotti, G. (2017). Open innovation and growth in IT sector. *Service Business, 11*(1), 45-68. doi:10.1007/s11628-015-0301-2

Bairi, J., & Murali Manohar, B. (2011). Critical success factors in gaining user customer satisfaction in outsourced IT services. *Journal of Enterprise Information Management, 24*(6), 475-493. doi:10.1108/17410391111166530

Barrad, S., Gagnon, S., & Valverde, R. (2020). An Analytics Architecture for Procurement. *International Journal of Information Technologies and Systems Approach (IJITSA), 13*(2), 73-98. doi:10.4018/IJITSA.2020070104

Baskerville, R. (1999). Investigating Information Systems with Action Research. *Communications of the Association for Information Systems, 2*(1), 19. doi:10.17705/1CAIS.00219

Baskerville, R., & Myers, M. D. (2004). Special Issue on Action Research in Information Systems: Making IS Research Relevant to Practice: Foreword. *MIS quarterly, 28*(3), 329-335. doi:10.2307/25148642

Chang, Y. B., & Gurbaxani, V. (2012). Information Technology Outsourcing, Knowledge Transfer, and Firm Productivity: An Empirical Analysis. *MIS Quarterly, 36*(4), 1043-1063. doi:10.2307/41703497

Chou, D. C., & Chou, A. Y. (2009). Information systems outsourcing life cycle and risks analysis. *Computer Standards & Interfaces, 31*(5), 1036-1043.

Coad, A., Daunfeldt, S.-O., Hölzl, W., Johansson, D., & Nightingale, P. (2014). High-growth firms: introduction to the special section. *Industrial and Corporate Change, 23*(1), 91-112. doi:10.1093/icc/dtt052

Coughlan, P., & Coghlan, D. (2002). Action research for operations management. *International Journal of Operations & Production Management, 22*(2), 220-240. doi:10.1108/01443570210417515

Dahlberg, T., & Lahdelma, P. (2007). IT Governance Maturity and IT Outsourcing Degree: An Exploratory Study. *Proceedings of the 40th Hawaii International Conference on System Sciences - 2007.* IEEE Xplore. doi:10.1109/HICSS.2007.306

Dallemule, L., & Davenport, T. H. (2017). What's Your Data Strategy? *Harward business review, 95*(3), 112-121.

DAMA. (2017). *DAMA DMBOK - Data Management Body of Knowledge.* (D. Henderson, Ed.) Technics Publications.

Delen, G. P., Reters, E. J., Verhoef, C., & van Vlijmen, S. F. (2016). Lessons from Dutch IT-outsourcing success and failure. *Science of Computer Programming, 130*, 37-68. doi:10.1016/j.scico.2016.04.001

Deloitte. (2021). *SaaS-enabled solutions deliver win-win for companies and customers*. Retrieved 3 2, 2021, from https://www2.deloitte.com/ch/en/pages/technology-media-and-telecommunications/articles/saas-enabled-solutions-deliver-win-win-for-companies-and-customers.html

Dhar, S., & Balakrishnan, B. (2006). Risks, Benefits, and Challenges in Global IT Outsourcing: Perspectives and Practices. *Journal of Global Information Management, 14*(3), 59-89. doi:10.4018/jgim.2006070104

Dibbern, J., & Heinzl, A. (2006). Selective Outsourcing of Information Systems in Small and Medium Sized Enterprises. In R. Hirschheim, A. Heinzl, & J. Dibbern, *Information Systems Outsourcing* (2nd ed., pp. 57-81). Berlin: Springer, Berlin, Heidelberg. doi:10.1007/978-3-540-34877-1

Dibbern, J., Goles, T., Hirschheim, R., & Jayatilaka, B. (2004). Information systems outsourcing: a survey and analysis of the literature. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 35*(4), 6-102. doi:10.1145/1035233.1035236

Eriksson, P., & Kovalainen, A. (2011). *Qualitative Methods in Business Research.* London, United Kingdom: SAGE Publications Ltd. doi:10.4135/9780857028044

Foerstl, K., Kirchoff, J. F., & Bals, L. (2016). Reshoring and insourcing: drivers and future research directions. *International Journal of Physical Distribution & Logistics Management, 46*(5), 492-515. doi:10.1108/IJPDLM-02-2015-0045

Gerbl, M., McIvor, R., & Humphreys, P. (2016). Making the business process outsourcing decision: why distance matters. *International Journal of Operations & Production Management, 36*(9), 1037-1064. doi:10.1108/IJOPM-04-2014-0192

Gualandris, J., Legenvre, H., & Kalchschmidt, M. (2018). Exploration and exploitation within supply networks: Examining purchasing ambidexterity and its multiple performance implications. *International Journal of Operations & Production Management, 38*(3), 667-689. doi:10.1108/IJOPM-03-2017-0162

Han, K., & Mithas, S. (2014). The Real Savings From IT Outsourcing. *MIT Sloan Management Review, 55*(2), 16.

Handfield, R., Seongkyoon, J., & Choi, T. (2019). Emerging procurement technology: data analytics and cognitive analytics. *International journal of physical distribution & logistics management, 49*(10), 972-1002. doi:10.1108/IJPDLM-11-2017-0348

Hirschheim, R., & Dibbern, J. (2006). Information Technology Outsourcing in the New Economy – An Introduction to the Outsourcing and Offshoring Landscape. In R. Hirschheim, A. Heinzl, & J. Dibbern, *Information Systems Outsourcing* (2nd ed., pp. 3-24). berlin: Springer, Berlin, Heidelberg. doi:10.1007/978-3-540-34877-1

Hirschheim, R., & Lacity, M. (2000). The Myths and Realities of Information Technology Insourcing. *Communications of the ACM, 43*(2), 99-107. doi:10.1145/328236.328112

Hodosi, G., & Lazar, R. (2013). How Do Critical Success Factors Contribute to a Successful IT Outsourcing: A Study of Large Multinational Companies. *Journal of Information Technology Theory and Application, 14*(1), 17-43.

Hätönen, J. (2010). Outsourcing and licensing strategies in small software firms: evolution of strategies and implications for firm growth, internationalisation and innovation. *Technology Analysis & Strategic Management, 22*(5), 609-630. doi:10.1080/09537325.2010.488066

Lacity, M. C., & Willcocks, L. P. (2009). Transaction Cost Economics Applied to IT Outsourcing: Findings and Critique. In M. C. Lacity, & L. P. Willcocks, *Information Systems and Outsourcing* (pp. 39-87). London, United Kingdom: Palgrave Macmillan. doi:10.1057/9780230594838

Lee, J.-N., Miranda, S. M., & Kim, Y.-M. (2004). IT Outsourcing Strategies: Universalistic, Contingency, and Configurational Explanations of Success. *Information Systems Research*, 110-131.

Limberakis, C. G. (2012). Spend Analysis: Lessons from the Best-in-Class. *Supply Chain Management Review*, 10-19.

Liyang, T., Zhiwei, N., Zhangjun, W., & Li, W. (2011). A Conceptual Framework for Business Intelligence as a Service (SaaS BI). *2011 Fourth International Conference on Intelligent*

Computation Technology and Automation (pp. 1025-1028). Shenzhen, China: IEEE. doi:10.1109/ICICTA.2011.541

Lopez-Garcia, P., & Puente, S. (2012). What makes a high-growth firm? A dynamic probit analysis using Spanish firm-level data. *Small Bus Econ*, 1029–1041.

Maržić, S., Krneta, P., & Pavlic, M. (2014). Spend analysis systems. *37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 1497-1502). Opatija, Croatia: IEEE. doi:10.1109/MIPRO.2014.6859803

Mclvor, R. (2008). What is the right outsourcing strategy for your process? *European Management Journal, 26*(1), 24-34. doi:10.1016/j.emj.2007.08.008

Mohr, J. J., Sengupta, S., & Slater, S. F. (2011). Mapping the outsourcing landscape. *Journal of Business Strategy*, 42-50.

Monczka, R. M., Handfield, R. B., Giunipero, L. C., & Patterson, J. L. (2016). *Purchasing and Supply Chain Management* (6th ed.). Boston, USA: Cengage Learning.

Mäkilä, T., Järvi, A., Rönkkö, M., & Nikkilä, J. (2010). How to Define Software-as-a-Service – An Empirical Study of Finnish SaaS Providers. In P. Tyrväinen, S. Jansen, & M. A. Cusumano (Ed.), *Conference proceedings ICSOB 2010* (pp. 114-124). Jyväskylä: Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-13633-7_10

Nadikattu, R. R. (2019). Data Warehouse Architecture – Leading the Next Generation Data Science. *International Journal of Computer Trends and Technology , 67*(9), 78-80. doi:10.2139/ssrn.3622840

Ngwenyama, O. K., & Bryson, N. (1999). Making the information systems outsourcing decision: A transaction cost approach to analyzing outsourcing decision problems. *European Journal of Operational Research, 115*(2), 351-367. doi:10.1016/S0377-2217(97)00171-9

Nicholls-Nixon, C. (2005). Rapid growth and high performance: The entrepreneur's "impossible dream?". *Academy of Management Executive, 19*(1), 77-89.

Novelli, F. (2011). A Simulation Study of the Interdependence of Scalability and Cannibalization in the Software Industry. *23rd European Modeling and Simulation Symposium (EMSS 2011)* (pp. 599-604). Rome, Italy: CAL-TEK SRL.

OECD. (2007). *Eurostat − OECD Manual on Business Demography Statistics.* Luxembourg: European Communities / OECD.

Oshri, I., Kotlarsky, J., & Willcocks, L. P. (2009). *The Handbook of Global Outsourcing and Offshoring* (3 ed.). London: Palgrave Macmillan. doi:10.1007/978-1-137-43744-0

Pandit, K., & Marmanis, H. (2008). *Spend Analysis: The window into strategic sourcing.* Fort Lauderdale, USA: J. Ross Publishing, Inc.

Qu, W. G., Oh, W., & Pinsonneault, A. (2010). The strategic value of IT insourcing: An IT-enabled business process perspective. *Journal of Strategic Information Systems, 19*(2), 96-108.

Salmela, H. (2008). Analysing business losses caused by information systems risk: a business process analysis approach. *Journal of Information Technology, 23*(3), 185-202. doi:10.1057/palgrave.jit.2000122

Sano, M. D. (2014). Business Intelligence as a Service: A New Approach to Manage Business Processes in the Cloud. *2014 IEEE 23rd International WETICE Conference* (pp. 155-160). Parma: IEEE. doi:10.1109/WETICE.2014.95.

Schniederjans, M. J., Schniederjans, A. M., & Schniederjans, D. G. (2005). *Outsourcing and Insourcing in an International Context.* New York, USA: M. E. Sharpe Inc.

Sharda, R., Delen, D., Turban, E., Aronson, J. E., Liang, T.-P., & King, D. (2015). *Business Intelligence and Analytics: Systems for Decision Support.* New Jersey: Pearson Education, Inc.

Sievo. (2021). Spend Classification with Sievo: User Manual.

Statista. (2021). Total size of the public cloud software as a service (SaaS) market from 2008 to 2020. Retrieved 3 2, 2021, from https://www.statista.com/statistics/510333/worldwide-public-cloud-software-as-a-service/

Sturgeon, T. J. (2021). Upgrading strategies for the digital economy. *Global Strategy Journal, 11*(1), 34-57.

Sun, W., Zhang, X., Guo, C. J., Sun, P., & Su, H. (2008). Software as a Service: Configuration and Customization Perspectives. *2008 IEEE Congress on Services Part II* (pp. 18-25). Beijing, China: IEEE. doi:10.1109/SERVICES-2.2008.29

Thouin, M. F., Hoffman, J. J., & Ford, E. W. (2009). IT outsourcing and firm-level performance: A transaction cost perspective. *Information & Management, 46*(8), 463-469. doi:10.1016/j.im.2009.08.006

van Gils, B. (2020). *Data Management: a gentle introduction* (1st ed.). 's-Hertogenbosch, The Netherlands: Van Haren Publishing.

Willcocks, L., Fitzgerald, G., & Feeny, D. (1995). Outsourcing IT: The strategic implications. *Long Range Planning, 28*(5), 59-70. doi:10.1016/0024-6301(95)00038-K

Yang, D.-H., Seongcheol, K., Nam, C., & Min, J.-W. (2007). Developing a decision model for business process outsourcing. *Computers & Operations Research, 34*(12), 3769-3778. doi:10.1016/j.cor.2006.01.012

Yen Ng, P., & Hamilton, R. T. (2016). Experiences of high-growth technology firms in Malaysia and New Zealand. *Technology Analysis & Strategic Management, 28*(8), 901-915. doi:10.1080/09537325.2016.1180355

APPENDICES

Appendix 1 - Participated Meetings and Interviews.

Date	Name	Context	Participants	Data Gathering Cycle
16.2.2021	Topic Confirmation	Introduction to the problem and context	CEO Sievo Oy	1st Data Gathering Cycle
23.2.2021	Implementation Function Overview	Introduction to data transformation in implementation organization	Implementation DM Team Lead	1st Data Gathering Cycle
5.3.2021	1st Steering Group Meeting	Introducing the project plan to the steering group	CFO, VP, Sales, VP, Implementation Function CEO Sievo Oy, VP, People & IT, CEO Sievo Inc.	Feedback/Planning
8.3.2021	New supplier consolidation – design presentation	Supplier Consolidation **training** (new)	Product Manager of Spend Module	2nd Data Gathering Cycle
9.3.2021	Unofficial Meeting	Talking about the topic in general	Multiple Continuous Service DM Specialists	1st Data Gathering Cycle
10.3.2021	Data Enrichment Operating Model Discussion	Talking about the project and helping to understand how to approach classification	Continuous Service DM Specialist and Classification Champion	1st Data Gathering Cycle
16.3.2021	PS Consolidation Task List Clarification	Understanding the nature of consolidation in implementation organization in-depth	Implementation DM Specialist and Consolidation Champion	2nd Data Gathering Cycle
17.3.2021	PS Classification Task List Clarification	Understanding the nature of consolidation in implementation organization in-depth	Implementation DM Team Lead	2nd Data Gathering Cycle
17.3.2021	Stakeholder Group Interview (consolidation)	Understanding the nature of technological development of consolidation. Clarifying what are the repetitive consolidation tasks in implementation and continuous	Implementation DM Team Lead, Product Manager of Spend Module, Continuous Service DM Specialist & Consolidation Champion, VP, Implementation Function	2nd Data Gathering Cycle

		service organizations.	Implementation DM Specialist and Consolidation Champion, Implementation Project Manager	
18.3.2021	Stakeholder Group Interview (classification)	Clarifying what are the repetitive classification tasks in implementation and continuous service organizations.	Implementation DM Team Lead, Product Manager of Spend Module, Customer Success Manager & Classification trainer, Continuous Service DM Specialist and Classification Champion, VP, Implementation Function	2nd Data Gathering Cycle
18.3.2021	March All Hands	Giving status update in company-wide monthly information meeting and invitation to participate with questions and comments	Every Sievo employee	Feedback
14.4.2021	Professional Services Monthly	Giving status update in implementation organization's monthly information meeting and invitation to participate with questions and comments	Every implementation function employee	Feedback
16.4.2021	2nd Steering Group Meeting	Giving a status update, confirming the following steps and the approach, and asking for comments for the work done so far	CFO, VP, Sales, VP, Implementation Function CEO Sievo Oy, VP, People & IT, CEO Sievo Inc.	Feedback/Planning
22.4.2021	Cost Analysis Planning 1st meeting	Planning how to calculate the cost of different scenarios	CFO	4th data gathering cycle
23.4.2021	1on1 In-depth Interview – DM Implementation Organization	Understanding the impact of executing different data	Implementation DM Team Lead	3rd data gathering cycle

		transformation tasks within different sce- narios		
26.4.2021	1on1 In-depth Interview – On- going DM Ser- vice Organiza- tion	Understanding the impact of executing the different data transformation tasks within different sce- narios	Continuous Service DM Team Lead	3rd data gathering cycle
26.4.2021	1on1 In-depth Interview – People & IT	Understanding the impact of executing the different data transformation tasks within different sce- narios	VP, People & IT	3rd data gathering cycle
27.4.2021	1on1 In-depth Interview – In- fosec	Understanding the impact of executing the different data transformation tasks within different sce- narios	Chief Infor- mation Security Officer	3rd data gathering cycle
27.4.2021	1on1 In-depth Interview – Sales	Understanding the impact of executing the different data transformation tasks within different sce- narios	VP, Sales	3rd data gathering cycle
27.4.2021	1on2 In-depth Interview - En- gineering	Understanding the impact of executing the different data transformation tasks within different sce- narios	Spend Product Manager, Spend Engi- neering Man- ager and Prod- uct Owner	3rd data gathering cycle
28.4.2021	1on1 In-depth Interview – Customer Or- ganisation	Understanding the impact of executing the different data transformation tasks within different sce- narios	CEO Sievo Inc.	3rd data gathering cycle
29.4.2021	1on1 In-depth Interview – Supplier Con- solidation	Understanding the impact of executing the different data transformation tasks within different sce- narios	Continuous Service DM Specialist & Consolidation Champion,	3rd data gathering cycle

30.4.2021	3rd Steering Group Meeting	Giving a status update, discussing cost analysis principles and supplier requests for information	CFO, VP, Implementation Function CEO Sievo Oy, VP, People & IT, CEO Sievo Inc.	Feedback/Planning
3.5.2021	1on1 In-depth Interview – Spend Classification	Understanding the impact of executing the different data transformation tasks within different scenarios	Continuous Service DM Specialist and Classification Champion	3rd data gathering cycle
14.5.2021	Cost Analysis Planning 2nd meeting	Planning how to calculate the cost of different scenarios	CFO	4th data gathering cycle
17.5.2021	Cost Analysis Planning 3rd meeting	Planning how to calculate the cost of different scenarios	CFO	4th data gathering cycle
18.5.2021	Cost Analysis Planning 4th meeting	Planning how to calculate the cost of different scenarios	CFO	4th data gathering cycle
20.5.2021	May All-Hands	Giving status update in company-wide monthly information meeting and invitation to participate with questions and comments	All the Sievo employees	Feedback
21.5.2021	4th Steering Group Meeting	Analysing the accuracy of established cost analysis.	CFO, VP, Sales, VP, Implementation Function CEO Sievo Oy, VP, People & IT,	Feedback/Planning
26.5.2021	Cost Analysis Planning 5th meeting	Making sure that quantitative analysis includes all the transaction costs	CFO	4th data gathering cycle
28.6.2021	5th Steering Group Meeting	Going through quantitative cost analysis and findings of qualitative analysis	CFO, VP, Implementation Function CEO Sievo Oy, CEO Sievo Inc.	Feedback/Planning

Appendix 2 – Interview Question Pool

Date:

Interviewee:

Title:

Function:

Efficiency

1. How would different scenarios impact the operating efficiency on task-level and managerial-level?
2. How process improvement, an increasing number of customers, and technological development are expected to impact repetitive workload?

Quality

3. How would you describe the quality level of current tasks?
4. How is the quality of repetitive tasks measured/codified?
5. How would different scenarios impact those levels?
6. What kind of direct communication with a client is required for executing the different tasks?
7. Which internal roles contribute to the classification and consolidation?

Required Expertise

8. What is the required expertise for conducting the tasks at a sufficient level? (technology, industry, customer-specific)? Does it variate between the tasks?
9. Do you think that we could obtain or lose crucial skills within different scenarios?

Impact on Customer Perception

10. How do clients perceive us currently?
11. How would that impression change within different scenarios?
12. How would the different scenarios impact the contracts that we currently have with our clients?

Relative Capability Position

13. Do we have a competitive advantage because we conduct the tasks in a superior or unique manner relative to our competitors?
    a. If yes, what would happen to the relative capability position within the different scenarios?

14. Does the relative capability exist only in certain tasks?
15. What are the critical skills or assets that enable us to perform in a superior manner?
16. Can we protect ourselves against the risks of transferring people and assets?

Contribution to competitive advantage (=key differentiator in the eyes of the customer)

17. Is the specific task critical to Sievo's competitive advantage directly or indirectly?
    a. If yes, how?
18. How would different scenarios impact our competitive advantage?

Required Management Capability:

19. What level of communication is required for executing the tasks? Is there a significant difference between the tasks?
20. How much is day-to-day management required?
21. How would the nature of managing change within different scenarios?
22. What kind of organizational efforts would the different scenarios increase and decrease?
23. Would some of the tasks be more feasible to be transferred than some other ones from a managerial perspective?

Technological Development Capability

24. How would you describe repetitive tasks' technological maturity?
25. Are there tasks that would not be technology-wise feasible to execute within the different scenarios?
26. How is the nature of the repetitive task expected to change in the near future?
27. What kind of technological efforts would the different scenarios require?
28. How would different scenarios impact technological development?
    a. Do you see that an external service provider could supply or support our technical development?
29. Do we have technologies that should not be accessible for an external operator?
30. Is there something else that Sievo should consider technology-wise?

Potential for Opportunism

31. What are the technological switching costs: How easily could the tasks be transferred to a subsidiary or external partner? Or back?) How those differentiate between different scenarios?
32. Are the repetitive task directly dependent on some other tasks? What kind of other tasks should be transferred along with the repetitive tasks?
33. How crucial are these tasks for Sievo's service provision? Do you see that there is potential for creating overdependence on the external service provider?

Security

34. What kind of technologies would we be able to transfer security-wise within different scenarios?
35. How would different scenarios impact our vulnerability and accessibility?
36. Would there be new emerging threats?
37. What kind of effort would different scenarios require security-wise?
38. How would different scenarios impact our contracts and arrangements with the clients?

Impact on company culture

39. What kind of HR efforts would different scenarios require? What are the main points to take into consideration when comparing insource vs. outsource and near-shore vs. offshore?
    a. Personnel
    b. Legal
    c. Culture
    d. Political
40. How would different scenarios impact company culture?
41. Would it be possible to guarantee the same performance from external employees than internal?
42. Would it be possible to offer the same benefits for all the employees in different scenarios? If not, what kind of consequences would that cause?

Appendix 3 – Detailed List of Data Transformation Tasks

Customer / Continuous Service Function			
Task	**Description**	**Customer Facing Task**	**Part of the Potential Project Scope**
1. Classification: Monthly Process: Product dimension	Most of the monthly classifications are done in the product dimension. The classification coverage must correspond agreed service level. For most of the clients, it is 95%.	x	x
2. Classification: Monthly Process: Other dimensions	In addition to product dimension, some clients also require classifying other dimensions, e.g., country, division, payment term, or supplier country. In those dimensions, 100% classification coverage is expected.	x	x
3. Classification: Other non-regular classification work	Ad-hoc tasks to reclassify spend without existing rules, e.g., existing client implements a new data source or increases its service level and is entitled to higher total classification coverage	x	x
4. Classification: Handling repetitive reclassification requests	Implementing written reclassification request received from a client	x	x
5. Classification: Mass row-level reclassification	Conducting a mass classification of hundreds of transactions by exploiting SQL procedures based on clients' structured feedback data.	x	
6. Classification Validation: Validating non-regular classification work	Classification validation is done in connection with ad-hoc classification work, not on a monthly basis. It is about looking at the top 20-30		x

	suppliers on each mid-level category of the taxonomy and checking, do these suppliers provide services/materials related to the category.		
7.1 Consolidation Validation: Validate automatic consolidation results	Validate that AI has consolidated correctly new monthly suppliers		x
7.2 In case of error: update from the existing suppliers the correct one	In case of AI consolidation error, choose the correct one from the other existing ones		x
7.3 In case of error: create a new supplier	In case of the correct supplier does not exist in the database, create a new one		x
8. Consolidation Validation: Handling repetitive reconsolidation requests	Correct existing supplier consolidations based on the client's feedback	x	x

Professional Services / Implementation Function			
Task	**Description**	**Customer Facing Task**	**Part of the Potential Project Scope**
1.1 Classification: Classify direct/material spend	Spend classification until 90% coverage of total spend, evenly distributed between direct and indirect in proportion to the amount of spend	x	x
1.2 Classification: Classify indirect/account spend	Spend classification until 90% coverage of total spend, evenly distributed between direct and indirect in proportion to the amount of spend	x	x
2. Classification: Handling repetitive reclassification requests	Implementing written reclassification request received from a client	x	x
3.1 Classification Validation: Category Validation	Looking at the top 20-30 suppliers on each mid-level category of the taxonomy and checking, do these suppliers provide services/materials related to the category		x
3.2 In case of error: correct classification	If the supplier is under the incorrect category, the correct category must be researched, and then the transaction reclassified		x
4.1 Classification Validation: Supplier Split validation	Looking at the top 100-200 suppliers by total spend and checking if there are small amounts of spend falling to unrelated tail categories due to incorrect account/material group bookings		x
4.2 In case of error: review classification model (not repetitive)	If there are lots of tail spend in unrelated categories, the classification model should be reviewed for understanding the root cause	x	

4.3 In case of error: correct classification	The tail spend should be reclassified to the correct category		x
5.1 Classification Validation: Material Split validation	Looking at the top 100-200 materials by total spend and checking if there are small amounts of spend falling to unrelated tail categories due to incorrect account/material group bookings		x
5.2 In case of error: review classification model (not repetitive)	If there are lots of tail spend in unrelated categories, the classification model should be reviewed for understanding the root cause	x	
5.3 In case of error: correct classification	The tails spend should be reclassified to the correct category		x
6.1 Classification Validation: Category depth validation	Checking if the classifications are done on the most detailed level of the taxonomy that data granularity permits.		x
6.2 In case of lack of depth: reclassify	If a significant portion of spend is classified only to a high-level category, one must check the spend and try to classify it on a deeper level of the taxonomy		x
7.1 Consolidation Validation: Validate top spend suppliers	Checking 90-95% of the total spend, around top 1000 vendors. Usually, 95% of spend covers 10-15% of suppliers		x
7.2 In case of error: update from the existing suppliers the correct one	In case of AI consolidation error, choose the correct one from the other existing ones		x
7.3 In case of error: create a new supplier	In case of the correct supplier does not exist in the database, create a new one		x
8. Consolidation Validation: Handling repetitive reconsolidation requests	Correct existing supplier consolidations based on the client's ad-hoc feedback	x	x

9.1 Consolidation Validation: Validate top clusters	AI sometimes consolidate vendors based on wrong data points, e.g., name of a city (all vendors including word 'Helsinki'), leading to incorrect consolidations. Therefore, additional validation is required by checking top supplier clusters		x
9.2 In case of error: delete and correct incorrect clustering	Delete and correct incorrect clustering done by AI		x