Defining use cases for Data Integration in a Digital Decoupling Integration Architecture at Olympus

Bachelor's Thesis

submitted by:

Simon Reinersmann

at:

NORDAKADEMIE Hochschule der Wirtschaft

1st Examiner	Prof. Dr. Michael Skall
2nd Examiner	Prof. Dr. Michael Schulz
Corporate supervisor	Stefan Tanck
Company	Olympus Europa SE & CO. KG
Matriculation number, centuria	10049, I19a
Address	Marienthaler Str. 69, 20535 Hamburg
Submission date	December 23rd 2022

Confidentiality Clause

This Bachelor's Thesis is based on internal, confidential data and information of the following company: Olympus Europa SE & CO. KG. This work may only be made available to the first and second examiner and authorised members of the board of examiners. Any publication and duplication of this Transfer Paper – even in part – is prohibited. An inspection of this work by third parties requires the expressed permission of the author and the company.

Note of Thanks

This thesis has been made possible through the help of many people.

First and foremost I would like to thank my corporate supervisor Stefan for his guidance and support during this thesis. His constructive feedback regarding the design of this research project and genuine interest in the results produced helped greatly. The commitment he had to supporting this project in a time of major changes in the Integration team and the Olympus IT, as well as during various business trips, deserves my utmost gratitude.

Furthermore, I would like to thank everyone who has proofread this thesis and all the people who contributed to this thesis. Special thanks are given to the following Olympus employees: Mark, Achim, Dennis, Marc, Christoph, John, and Anne.

Finishing this thesis marks the end of my bachelor's programme. I have met many people who have positively influenced me and helped me during my studies. The most important person is my partner Emmy. I am supremely beholden for her unconditional support both morally and directly. Furthermore, and in no particular order, I would like to thank Christine, Christopher, Felix, Finn, Jonas, Kira, Leif, Lina, Lobana, Özlem, Silas, and Timo for all the good times we had during our studies and the good friendships we developed.

Finally, I would like to thank my parents Heike and Dieter and my sister Ricarda for their continuous support. As the greatest influence in my life, I would like to thank my family for everything.

Contents

List of Figures

List of Tables

Abbreviations

ADM Architecture Development Method.

API Application Programming Interface.

B2B Business to Business.

B2C Business to Customer.

BI Business Intelligence.

CDC Change Data Capture.

CMS Content Management System.

CPI Cloud Platform Integration.

CRM Customer Relationship Management.

DCX Digital Customer Experience.

DIH Digital Integration Hub.

DnA Data and Analytics.

E2E End-to-End.

ELT Extract, Load, Transform.

EMEA Europe, Middle East, and Africa.

ER entity-relationship.

ERP Enterprise-Resource-Planning.

ETL Extract, Transform, Lead.

GTM Go to Market.

HIP Hybrid Integration Platform.

IaaS Infrastructure as a Service.

iPaaS Integration Platform as a Service.

ISA-M Integration Solution Advisory Methodology.

MBC Manufacturing Business Centre.

MDG Master Data Governance.

MDM Master Data Management.

med-tech medical technology.

OAO	Olympus Asia and Oceania.
OCA	Olympus Corporation of the Americas.
OCN	Olympus China.
OEKG	Olympus Europa SE & CO. KG.
OLAP	Online Analytical Processing.
OLTP	Online Transaction Processing.
OT	Olympus Tokyo.
PaaS	Platform as a Service.
SaaS	Software as a Service.
SBC	Sales Business Centre.
SCM	Supply Chain Management.
SOA	Service-Oriented Architecture.
TOGAF	The Open Group Architecture Framework.

1 Introduction

The modern world is continuously becoming more reliant on data. In the field of information technology, collecting and utilising data has become imperative for fields ranging from business models and entire companies built on the value of data to newly emerged research topics. This trends can be seen when reviewing the most popular search terms and most cited articles on digital libraries.[1] The importance to enterprises is displayed in analyst forecasts of relevant trends for companies in the coming years. Here, many trends are highly dependent on data and Data Integration.[2] Furthermore, proven technologies leveraging data play an increasingly important role for businesses capitalising on data to create new business models.[3]

Therefore, data has become an essential part of the modern world and businesses operating in it.[4] However, actually leveraging data to gain knowledge, or add value is complex and depends on the integration of the data.[5] Full integration of data often proves to be too complicated, resource intensive and rigid for the use cases defined for enterprise integration.[6] The reasons for this lie in the nature of the high frequency of developments in systems involved in integration, the high complexity of delivering integrions, and an increasing demand for integrations.[7] Several integration approaches have been designed to offer solutions to these problems.

The term *Digital Decoupling* describes a system for integration, which in turn determines an Integration Architecture for information systems.[8] The main goal of using this architecture in conjunction with a Digital Decoupling system is to enable legacy systems and new computing technologies to work in conjunction with each other and to decouple these systems to decrease the time to market, lower efforts and costs.[9] In this case, decoupling describes the approach of eliminating a direct coupling of source systems to target systems and instead leveraging the Digital Decoupling System as a middle ware platform.[10] Therefore, the system aims to solve two fundamental problems of enterprise integration: Having a high degree of integration whilst coupling systems loosely and enabling asynchronous communications between system.[11]

[1] Association for Computing Machinery (2022); Institute of Electrical and Electronics Engineers (2022).
[2] Groombridge (2022):pp. 9, 11, 14–17, 23, 28.
[3] Raj et al. (2020):p. 9.
[4] Fang (2015):p. 820; Sorokine et al. (2016):p. 34; Cuzzocrea and Ciancarini (2021):p. 6.
[5] Pezzini and Thoo (2020):p. 1.
[6] Mertens (2013):p. 19.
[7] Pezzini (2018):pp. 10–11.
[8] Zeier, Wagenknecht, and Kleber (2021):p. I.
[9] Zeier, Wagenknecht, and Kleber (2021):p. I; Accenture Banking and Israel (2019).
[10] Zeier, Wagenknecht, and Kleber (2021):p. 2.
[11] Hohpe and Woolf (2004):p. XIX.

1.1 Outline of the problem and goal of the thesis

From a global perspective, Olympus' IT landscape has grown to be highly fragmented and regionalised. This fragmentation has developed over time as the company expanded from Japan into new markets. As the new markets were captured by building regional subsidiaries which had considerable leeway for the way in which the business was conducted, there was no overarching structure for the organisation, and thus also the IT. With Olympus' goal to become a global leading medical technology (med-tech) company, this has changed drastically. This global transformation requires existing systems within these subsidiaries to be adapted to meet the business requirements of a global organisation. In detail, this affects the integration of numerous local systems that are hosting similar data for their region into global solutions. This scattered nature of the data led to Olympus developing the concept of the Global Integration Hub, which aims to resolve this problem by offering a global single point of truth for Data Integration. However, this system has so far neither been implemented, nor has the usefulness to Olympus been clearly evaluated.

The goal of this thesis is to define use cases for Data Integration leveraging Digital Decoupling. To define these use cases and assess their impact on Olympus as a global enterprise, the following research questions will be answered in this thesis:

RQ1: How can different Integration Architectures like Digital Decoupling support Olympus' effort to unify Data Integration globally?

RQ2: What are concrete use cases for Data Integration in the Global Integration Hub and how do these use cases benefit from the proposed Integration Architecture compared to other approaches?

The knowledge gained in this thesis is intended to serve as guidance to Olympus in deciding which use cases can benefit the most from the proposed Integration Architecture. Deficiencies in the approach to Data Integration using the Global Integration Hub will reveal possible improvements based on the best practises outlined in the specialist literature. Additionally, this critical analysis will give an insight into use cases where Data Integration using the Digital Decoupling Integration Architecture would pose drawbacks to other methods available to the Olympus Integration team.

1.2 Course of the thesis

This thesis is built in four parts: Starting with chapter 2 where core concepts of enterprise Data Integration are introduced. These concepts form the basis for more abstract models introduced in later chapters. In the first section, Data Integration is defined by demarking different types of enterprise integration. Building on this, the concepts of data warehouses and data lakes are introduced as manifestations of enterprise integration leading into the Extract, Transform, Lead (ETL) and Extract, Load, Transform (ELT) processes as means of batch Data Integration. Following this, Application Programming Interfaces (APIs) will be described, with a focus on web based API.

Chapter 3 builds on these foundational elements of enterprise integration by introducing Enterprise Architectures. For this, proven principles of this field like the Architecture Development Method (ADM) model of The Open Group Architecture Framework (TOGAF) are described. Building on this, the term Digital Decoupling Integration Architecture is developed. Furthermore, this Integration Architecture is analysed in detail, examining each component utilising the concepts introduced in chapter 2. Adding to this, the Digital Integration Hub (DIH) is analysed as a different Integration Architecture supporting the loose coupling of systems. Both chapters outlined so far are based on a qualitative literature review, meaning that the knowledge presented was obtained by carefully selecting and reviewing a variety of sources published in journals or books on these topics.

The theoretical knowledge gained is then applied to the enterprise integration landscape of Olympus in chapter 4. For this, the global organisation of Olympus is laid out and the effects of this organisation of the company on the IT landscape is reviewed. Hereinafter, the Olympus Global Integration Hub is reviewed and critically compared with the models evaluated in chapter 3. For this, the needs of the business will be defined.

The use cases defined in this thesis are then based on the potential of the Integration Architecture and the actual capabilities of the architecture as evaluated in chapters 3 and 4 respectively. For this, chapter 5 will determine use cases for Data Integration in the Digital Decoupling Integration Architecture for three selected initiatives at Olympus. The initiatives are concerned with different domains of Data Integration.

Finally, this thesis will be concluded by answering the research questions, proposing an outlook and pointing out opportunities for further research in chapter 6.

2 Fundamentals of Data Integration in an enterprise

Mertens defines Enterprise Integration as connecting singular elements to a holistic system.[1] Integration of different components and systems is considered to be one of the core capabilities that determine the success of an Enterprise-Resource-Planning (ERP) system.[2] Integration also marks one of the key paradigm shifts in ERP systems that occurred at the start of this century. This paradigm shift saw ERP systems transition from monolithic on-premise systems to a network of interconnected systems, based on on-premise and cloud hosting strategies.[3] Data Integration is one of the key concepts in Enterprise Integration, that enables this network. Especially with the advent of technologies that leverage big data, integration of different data sources has become a higher priority and more complex in nature due to heterogeneous data.[4]

The focus of this chapter lies on explaining basics of enterprise Data Integration building the foundation for succeeding chapters to analyse decoupling in Integration Architectures. For this, section 2.1 will depict principals of interrogations and differentiate the different types of integrations. Out of this, the term Data Integration will arise, which will be described in detail in section 2.2. For this a theoretical concept of Data Integration will be established subsequent, and the subsections will cover Data Warehouse (2.2.1) and Data Lakes (2.2.2) as practical implementations for Data Integration. This will lead to the topic of data extraction in section 2.3. Finally, the term API will be explained along with different perspectives on APIs (2.4.1), the role of APIs in Enterprise Integration (2.4.2), and the different audiences inside and outside an enterprise for APIs (2.4.3).

2.1 Types of Enterprise Integration

The nature of Enterprise Integration is to connect systems and modules that generally are heterogeneous.[5] This variety of integration tasks leads to a pool of structurally different kinds of integrations. This makes a clear differentiation and categorisation of the types of Enterprise Integrations difficult.

[1]Mertens (2013):p. 13.
[2]Chang (2004):p. 5; Shaul and Tauber (2013):p. 55:3.
[3]Shaul and Tauber (2013):p. 55:7.
[4]Kadadi et al. (2014):pp. 38-39.
[5]Hohpe and Woolf (2004):p. 2-3.

The specialist literature and companies that are in the industry define different ways to categorise Enterprise Application Integrations. For most of these categorisation efforts, some technical aspect of a given integration is used for this differentiation. A categorisation model that is often referenced was developed by Mertens[6].

Mertens provides five different dimensions for categorising integrations, which are lied out in table 2.1. One dimension is the integration range. It differentiates between integration of systems that are in a single business, and integration of systems that are distributed across multiple businesses.[7] However, this differentiation does not explicitly cover subsidiaries, leading to some uncertainties in this differentiation. This may however be by design, as the businesses in these relationships tend to have a varying degree of shared information systems. Alternatively, this categorisation could be done by differentiating between internal and external integrations and further splitting up the external integrations into supplier integration and customer integration.[8] This categorisation is in line with the structure of companies, their supply-chains, and most ERP systems, where a separation between customers and suppliers, or debtors and creditors is common practise.

Integration dimension	Possible manifestations
Integration object	Data Integration, Function integration, Process integration, Method integration, Application integration
Integration direction	Horizontal integration, Vertical integration
Integration Range	Section integration, Process transcending integration, Intra company integration, Inter company integration
Degree of automation	Fully automated, Partially automated
Integration Time	Batch job integration, (Near-)real-time

Table 2.1: Different kinds of integrations according to Mertens (Own table based on Mertens (2013):p. 14)

Another way of differentiating integration types is to classify them by integration direction. This leads to a differentiation in horizontal and vertical integration. The former describes integration in a value-adding stream, whilst the latter covers the integration of strategic and operational systems.[9] Furthermore, integration can be differentiated by the degree of automation, which

[6]Mertens (2013).
[7]Mertens (2013):pp. 19–20.
[8]Koufteros, Vonderembse, and Jayaram (2005):pp. 100–106.
[9]Mertens (2013):p. 18.

ranges from fully automatic to various forms of human-machine interaction.[10] This also affects the integration time, which can be real-time or near-real-time, if the integration is fully automatic. Alternatively, integrations can be enabled through batch jobs.[11]

The most common differentiation of integrations, which is also the most compatible to other categorisation systems, is based on the integration object. In this case, objects can be data, functions, processes, methodologies, or applications.[12] This last categorisation is similar to other sources in the specialist literature and the SAP Integration Solution Advisory Methodology (ISA-M) framework[13].

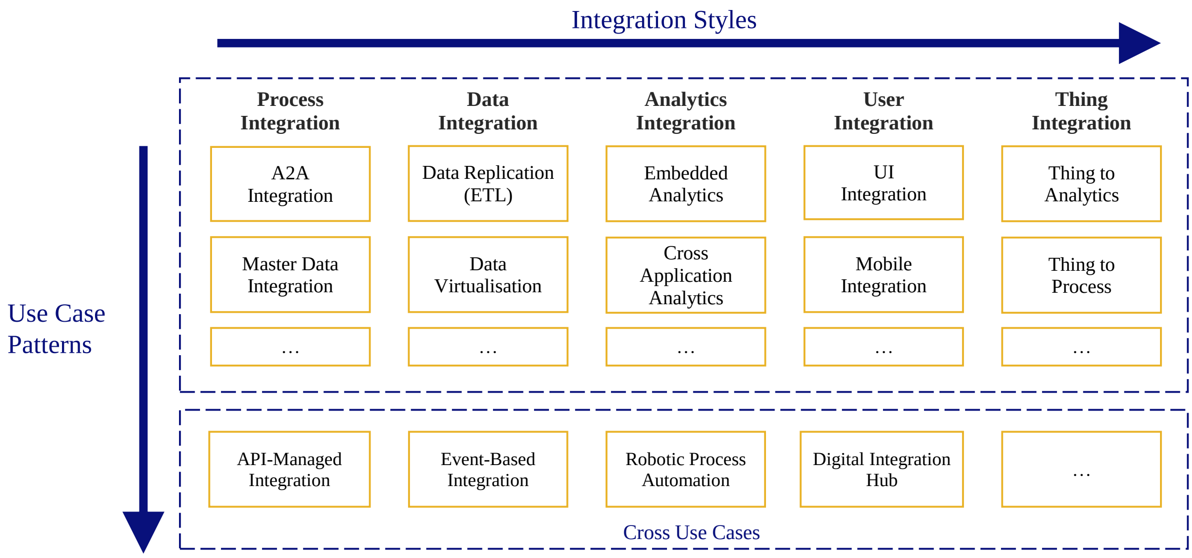


Figure 2.1: A selection of the integration styles and use case patterns in the ISA-M framework *(Own representation based on Allgaier (2022):p. 27)*

Figure 2.1 shows that, the framework also classifies integration styles by process- and Data Integration, among others.[14] However, the other integration styles differ from the integration objects defined by Mertens. Since February 2021, SAP also defines analytics integration, which in Mertens' system would be covered by Data Integration.[15] Furthermore, the ISA-M framework covers user- and thing integration, which cannot be directly mapped to any of Mertens'

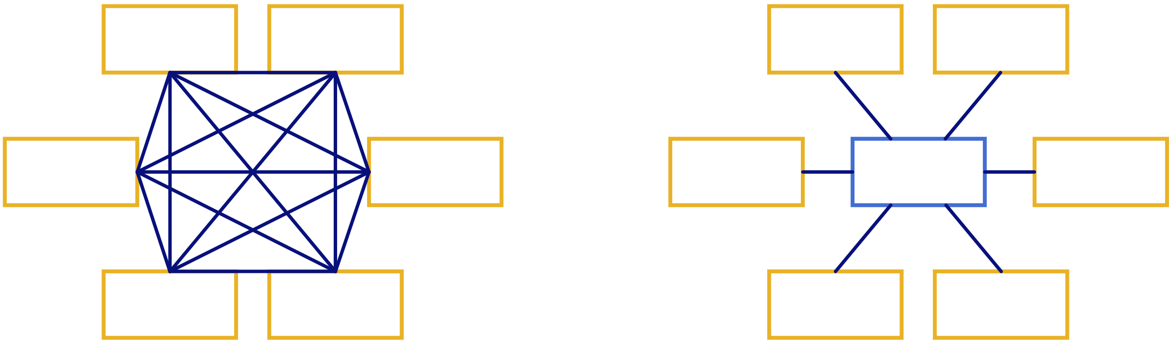
[10]Mertens (2013):pp. 21–23.
[11]Mertens (2013):p. 23.
[12]Mertens (2013):pp. 13–14.
[13]Allgaier (2022); SAP is short for the German "Systemanalyse Programmentwicklung". However, as this acronym fades out, it is not listed as an abbreviation. SAP is a leading producer of enterprise software like ERP systems. This has two effects on the sources provided by SAP. The first is that the word of SAP has gravity in the industry. Therefore, the standards, frameworks, and technologies developed by SAP have an impact on the industry. The second effect is that these sources mostly relate to products or services of SAP.
[14]Allgaier (2022):p. 24.
[15]Allgaier (2022):p. 9.

dimensions.[16] A reason for this variation between these sources may be caused by SAP altering the integration framework to better align with their product portfolio. Another reason could be that the ISA-M framework is more in line with modern integration patterns, as it is updated frequently. The change log provided within the framework shows that changes have been made to align the framework with the SAP naming conventions and new products.[17] For this reason, the ISA-M framework aligns very well with SAP products. However, SAP constructed the Framework to be vendor- and product-agnostic.[18] Still, other sources in basic literature and current journal articles also have to be considered, to obtain a holistic view of this topic. The terms process integration and Data Integration, however, are universally used in different pieces of literature by experts and companies alike.[19] These two types of integration also seem to be the most persistent in multiple decades of research.

This is particularly interesting, as sources offering practical guidance advice on the implementation of different integration methods also use this differentiation. However, instead of strictly segmenting integration efforts according to these categories a unified approach across the company and integration domains is recommended, especially regarding data- and application integration.[20]

Another way of differentiating Enterprise Integration is by analysing whether a central hub is being used, or if the integration connects systems directly with one another. The latter system is called a point-to-point or peer-to-peer connection and is shown in fig. 2.2a.[21] The central point in a hub-and-spoke architecture is the message broker, which in itself can be considered an integration paradigm.[22] In fig. 2.2b this model is shown with the central message broker in light blue. The integration through a central hub does not limit the range of Data Integration capabilities that can be provided. According to Gartner's Thoo and Friedman, multiple Data Integration styles should be leveraged in a hub-style system to take advantage of each of their strengths. Thereby, a hub style Data Integration should lead to a higher reusability of integrations, as fewer integrations are needed compared to a point-to-point integration.[23]

[16] Allgaier (2022):pp. 24–27.
[17] Allgaier (2022):p. 8-9.
[18] Allgaier (2022):p. 4.
[19] Hohpe and Woolf (2004):pp. 5–6; Funk et al. (2010):p. 87-91; Mertens (2013):pp. 13–14; Kiwon et al. (2021):p. 72; Allgaier (2022):p. 24.
[20] Thoo and Friedman (2017):pp. 7, 9.
[21] Kiwon et al. (2021):p. 52.
[22] Hohpe and Woolf (2004):p. 324.
[23] Thoo and Friedman (2017):pp. 1, 4.



(a) Point-to-point integrations

(b) Hub-and-spoke integrations

Figure 2.2: Comparison of point-to-point and hub-and-spoke integrations using a central message broker *(Own representation based on Hohpe and Woolf (2004):pp. 322, 324)*

The main difficulty in deploying a hub-and-spoke model, as it is also called, lies in defining a common ground for data because different systems provide different interfaces and message types. Moreover, when a variety of real world production systems is integrated at a central point, differences in the data, or the way the data is modelled are bound to occur. However, managing this proves to be easier when connecting numerous enterprise systems to a central hub than managing the number of connections required for point-to-point connections.[24] This problem is not unique to Enterprise Integration, but is a common issue of networks of systems. An analogy to this can be found in basic networking, where switches or routers are used to act as central connection points. With hubs, the number of connections that need to be made can be equal to the number of systems that are connected, if only one hub is used and systems do not have any connections besides the connections to the hub. With point-to-point connections the *number of connections* c will rise exponentially with the *number of systems* n as in $c = \frac{n*(n-1)}{2}$, if every point-to-point connection is made exactly once. Both of these cases can be seen in fig. 2.2. However, even if only some point-to-point connections are made, the point of diminishing returns will set in with a relatively low number of systems.

[24]Funk et al. (2010):pp. 88–89.

2.2 Enterprise Data Integration

Data Integration has been and continues to be a highly relevant and researched topic. It has strong links to Enterprise Integration and Business Intelligence (BI), which is concerned with retrieving information from data.[25] The age of big data and data science has especially fuelled this topic. However, the underlying problem, that Data Integration aims to solve, has stayed the same. This problem of extracting and transforming data from heterogeneous sources into a common data model has been covered in different ways.[26] From a theoretical perspective, a "*Data Integration system* I [is formalised] in terms of a triple $\langle G, S, M \rangle$, where G is the *global* schema [...], S is the *source* schema [...], M is the *mapping* between S and G, constituted by a set of *assertions* of the forms $q_S \rightsquigarrow q_G$, $q_G \rightsquigarrow q_S$[...]"[27] This formalisation displays the fundamentals of Data Integration very clearly. It shows the three main components of a Data Integration system: A source data schema, a global data schema and a mapping between the two, that enables the transformation from one schema to the other and back. The components of the source data schema and the mapping may exist multiple times, since multiple sources with different data schemata will call for more than one mapping logic. There may also be multiple steps to the mapping logic. However, this formalisation also hides the complexity that is hidden in developing a sound mapping, which is the main complexity in Data Integration.[28]

Data Integration typically uses some form of replication of the data to integrate it into other systems.[29] However, there is also the possibility of manipulating the data. This is expressed in the difference between ETL and ELT which is covered further in section 2.3.[30]

2.2.1 Data Warehouse

A common system for Data Integration is the Data Warehouse. A Data Warehouse typically stores data from different sources in a unified schema that is optimised for BI applications.[31] This has a few indications for Data Warehouses:

[25] Sreemathy et al. (2020):p. 1446.
[26] Lenzerini (2002):p. 235.
[27] Lenzerini (2002):p. 234.
[28] Kadadi et al. (2014):p. 39.
[29] Kiwon et al. (2021):p. 76.
[30] Sreemathy et al. (2020):p. 1444.
[31] Sreemathy et al. (2020):p. 1444; Kiwon et al. (2021):p. 76.

1. Data needs to be aggregated from other systems and transformed into the schema of the Data Warehouse. This makes the integration of incomplete data difficult, which in turn makes integration of streaming data sources more complicated.

2. The transformation into a single data model eases the access to the transformed data, but also limits the information retrieval, to the transformed data. This means that any data that got changed or removed during the transformation cannot be analysed.

Traditionally, Data Warehouses rely on database systems built for Online Analytical Processing (OLAP). These systems are designed for queries, typically used by BI. Therefore, these systems differ from Online Transaction Processing (OLTP) database system, that are conventionally the production database systems of a company. Data in OLAP systems is typically arranged in multiple dimensions and stored in a star schema, where the central fact table holds the primary key for the dimension, the secondary keys for the associated tables, and sometimes further data for each fact. Less often, OLAP systems can also store data in the snowflake schema, which is a more normalised form of the star schema. The snowflake schema is differentiated by secondary tables that also store secondary keys. When visualised in an entity-relationship (ER) model with the fact table in the middle and the other tables laid out in a circle around the fact table, the names of each schema become clear.[32] Typically, data arranged in a star schema is in the second normal form, while the snowflake schema typically arranges data in the third normal form. With the onset of in-memory column-based databases, the need for the separation of OLTP-style and OLAP-style systems was questioned by prominent figures like Hasso Plattner. Furthermore, Plattner proposes that Data Warehouses are a compromise and that in-memory column-based databases cloud benefit OLTP platforms and bring parity, therefore leading to OLTP systems assimilating to OLAP systems.[33] However, this proclaimed paradigm-shift will also have been linked to the development of SAP's S/4 HANA, which utilizes this technology. Therefore, this source has to be considered tentatively. When comparing the proposed benefit with adoption rate of products that utilize this technology like SAP S/4 HANA, it becomes clear that factors like total cost of ownership or complexity of the system in migration and operation outweigh these proposed benefits.[34]

Current trends in Data Integration concern the growing number of data sources and formats and the trend towards leveraging public- and unstructured data, for example, in big data networks.[35]

[32]Chaudhuri and Dayal (1997):pp. 66, 68–70.
[33]Plattner (2009):pp. 1, 5–6.
[34]Scott (2021):p. 1.
[35]Kiwon et al. (2021):p. 77.

With the advent of these trends, new problems emerged, and some existing problems got worse. For example, as the size and number of data sets grow, the access time to the data and the time for data manipulation plays a crucial role. This is due to a combination of hard- and software limitations[36] but also due to a lack of know-how in companies.[37] An early solution to this problem was to have users choose how to transform the raw data upon the use of the data to be integrated. This allows for great flexibility in the creation of reports, whilst at the same time having little upfront work in integrating and therefore transforming the data.[38] This paradigm shift has led some authors to call the Data Lake that will be described in section 2.2.2 the evolution of a typical Data Warehouse,[39] whilst other authors were proclaiming the death of traditional Data Warehouses.[40] Other authors still propose a combination of both techniques into a singular data modelling pipeline.[41]

2.2.2 Data Lake

This fundamental idea of storing the raw data loaded from the source systems before being transformed in any way has evolved into a system known as the Data Lake.[42] This makes Data Lakes suitable for many different applications in businesses and induces a large amount of research on this topic. Data Lakes typically store high volumes of heterogeneous data that can be structured, semi-structured, or unstructured. Due to this, the main challenge also lies in the information retrieval process associated with deploying these systems. Therefore, different methods have emerged that aim to make this process less computationally expensive. An example of one such method would be data set proximity mining, which leverages the metadata stored in the Data Lake. In general, Data Lakes rely on metadata to be able to retrieve information from the data stored.[43]

In what is considered to be a landmark paper on this topic, Fang[44] describes the underlying idea and advantages of Data Lakes. These include fundamental differences to Data Warehouses like

[36] These limitations were more pronounced before a more widespread adoption of in-memory databases that increase data throughput whilst greatly decreasing access latencies. However, a limitation of this type of database is the typically smaller sizes available to the databases compared to conventional systems that make use of disk arrays. However, this obstacle may be overcome with more modern server platforms and software design solutions like space-time trade-offs. (See Zhang et al. (2015):p. 1921.)
[37] Kadadi et al. (2014):p. 39.
[38] Knoblock and Szekely (2013):p. 28-29.
[39] Fang (2015):p. 820.
[40] Abelló (2015):p. 35.
[41] Raj et al. (2020):p. 18.
[42] Raj et al. (2020):p. 18.
[43] Alserafi et al. (2017):pp. 1–2, 4.
[44] Fang (2015).

schema on read. This also describes the use of raw data, but explicitly formulates, that the data schema is created, when the data is analysed.[45] Compared to schema on write, which is often used in Data Warehouses, this offers some distinct advantages:

- The degree of flexibility provided to the data scientist is highest, as the format of the data loaded is of no concern to the Data Lake.[46] This also implies, that unforeseen discoveries can be made, that may have not been possible without access to the raw data. This is because data transformation may lead to significant loss of data.[47]
- The use of raw data enables the support of new data-types. This also speeds up load times.[48]

Furthermore, the paper goes into what has to be done to be able to take advantage of a Data Lake. One of the key findings is that a Data Lake does not eliminate the core problem of Data Integration, which is the transformation of data to discover information. The Data Lake can be considered as a kind of IT infrastructure, that needs to be utilised by other tools for actual information discovery and needs to be clearly understood by the business users.[49] Therefore, it is value enabling, not value adding.

However, there are hurdles that Data Lakes face due to their design of incorporating raw data from a variety of sources. Specialist literature describes the following problems of Data Lakes whilst acknowledging the value of being able to perform data analysis on raw data, without loosing potential input parameters through transformation:

- The governance of data quality is difficult, but crucial to the value of the data lake. When the data quality is allowed to degrade too much, the Data Lake's open structure becomes a major problem and the Data Lake turns into a swamp.[50]
- Adding to the difficulty of keeping a Data Lake's data quality high is the fact that the variety of sources Data Lakes are connected to means that the frequency of data ingestion into the Data Lake is highly irregular.[51] This can pose significant problems for avoiding and identifying duplicate data and for the consistency of data stored in the Data Lake.

[45] Fang (2015):pp. 820–821.
[46] Fang (2015):p. 821.
[47] Zagan and Danubianu (2021):p. 1.
[48] Fang (2015):p. 821.
[49] Fang (2015):p. 824.
[50] Abelló (2015):p. 35.
[51] Zagan and Danubianu (2021):p. 4.

- Generally, when Data Lakes are deployed they cannot effectively hold all the data, that might be available and interesting for knowledge discovery.[52]

Data Lakes are a somewhat new technology in the realm of enterprise software. Many deployments of such software now tend to favour some form of licensed service in the form of Infrastructure as a Service (IaaS), Platform as a Service (PaaS), or Software as a Service (SaaS), instead of licensing and self-hosting these applications. Specifically for Data Lakes, there are advantages to this approach, as the data stored will be of high value to the company.[53] Furthermore, Data Lakes leverage low-cost storage for the characteristic high amounts of data.[54] For these reasons and due to their usefulness, Data Lake products are a staple of cloud computing providers.[55] However, these problems are outside the scope of this bachelor's thesis.

2.3 Data extraction

A prerequisite for Data Integration is that data in the source system can be accessed or extracted out of the system to integrate it into the target system. The specialist literature proposes different methods of Data Integration. These different approaches can be differentiated by whether the data is being physically moved or manipulated in situ. When physically moving the data to another system, one of the most common methods is the ETL process.[56] Manipulating the data in situ will be outside the scope of this bachelor's thesis. Furthermore, data pipelines for Data Integration can be categorised into ETL and ELT systems.[57]

As the name ETL implies, there are three steps in this process. These three steps are to extract the data from the source system, transform the data from the source format to the destination format, and load the transformed data into the target system.[58]

Extraction is the first step in the ETL process. This step is focused on retrieving structured or semi-structured data from various source systems and loading the data into a database in the ETL system. After this step, the data will still be in the data schema of the source system.[59] The extraction can be performed either fully or in part. Partial extractions may be triggered by an update notification of the source system.[60]

[52] Zagan and Danubianu (2021):p. 4.
[53] Zagan and Danubianu (2021):p. 2-3.
[54] Fang (2015):p. 820.
[55] Amazon Web Systems (2022); Google (2022); Micrisoft (2022).
[56] Sreemathy et al. (2020):p. 1444.
[57] Raj et al. (2020):p. 14.
[58] Prema and Pethalakshmi (2013):p. 430; Sreemathy et al. (2020):p. 1446.
[59] Sivabalan and Minu (2021):p. 2.
[60] Sreemathy et al. (2020):p. 1447.

Transformation marks the step in which the data that are now in the ETL system will be transformed. This means that the data will be cleaned, so that any errors in the data or missing data can be found and potentially fixed. Data will also be transformed to conform to the data schema desired for the target system.[61] Some data may not need to be transformed. This data is called Direct move or pass through data.[62] However, using ETL systems simply for extraction and loading could be a waste of resources, as the main value add of these systems lies within the transformation.[63] The actual transformation of the data from a source to a target schema has been covered in detail in section 2.2.

Load describes the actual transmission of transformed data from the ETL system to the target system. It is the final step in the ETL process. The loading can be performed as a full refresh, incremental load, or initial load based on the integration time[64].[65]

Through the design as a standalone system, ETL systems are particularly well suited to performing complex transformations on data sets of medium to small sizes. In complement, ELT systems excel at performing less complex transformations on larger sets of structured and unstructured data.[66] This is due to the design of these systems. In ELT processes, the target system is tasked with transforming the data.[67] This allows the server on which the database is running to handle the task of transforming the data. This eliminates the need for a powerful ETL system and can increase the transformation of the data, as typically database servers have higher performance compared to ETL servers.[68]

2.4 Application Programming Interface

The meaning of the acronym API is Application Programming Interface. As the name suggests, APIs are used to exchange data and information between two different applications, or more general systems. For this, they rely on a predefined syntax that defines the standard by which the information or data is to be exchanged. Therefore, APIs on their own, are neither systems nor databases. They are just the piece of the system that enables other applications to access a system's data. This nature is described in three core concepts of APIs: Modularity, interoperability, encapsulation. The use cases for APIs are very broad, as they are used among other things in

[61] Kabiri and Dalila (2013):p. 2.
[62] Sreemathy et al. (2020):p. 1447.
[63] Kimball and Caserta (2004):p. 113.
[64] Table 2.1.
[65] Sreemathy et al. (2020):p. 1447.
[66] Sivabalan and Minu (2021):p. 5.
[67] Sreemathy et al. (2020):p. 1445.
[68] Sivabalan and Minu (2021):pp. 3–4.

frameworks, software libraries, operating system, file systems, and drivers.[69] However, for the scope of this thesis only network-based or web-APIs are to be analysed, Enterprise Integration is mainly focused on these.

2.4.1 Perspectives on APIs

The term API can mean different things to different people. This can be attributed mainly to two different points of view on the topic. The first point of view is the technical one, where an API simply is the technical interface between systems. The second point of view frames the API as a digital product that enables the data access between systems. Both points of view are legitimate, but the technical implications have to be understood before implementing APIs. Out of these points of view, different perspectives derive: One is the Software Architect, to whom APIs are a technical tool for system integration. Another is a Product Owner, or Project Manager, who views APIs as a digital product. This perspective also coined the term API as a Service. The last perspective is that of the (IT-)management that views APIs as a way to connect systems, thereby integrating systems along a value chain.[70] Perspectives based on APIs as data products have had a particular increase in popularity among other trends that concern Data Integration.[71]

These perspectives can culminate in what is known as a Service-Oriented Architecture (SOA). This model was particularly popular in the decade from 2000 to 2010. In the case of the SOA the word architecture can have two distinct meanings. The first is a pattern that describes the interactions of systems on a high level. The second is a concrete development approach, that describes how modules of a system are to be developed. The term service has no concise definition in this context. Briefly, a service can be described as a set of rules that define the semantics with which it operates. This nature of this service is clear in the value proposition and the visibility of the service. However, the flexibility is unclear. Where the business perspective desires flexibility of the service to adapt to changing markets, the technical perspective wants to create a reusable service, that is long-lasting.[72]

2.4.2 APIs in Enterprise Integration

Enterprise Integrations are based on what Frank, Strugholtz, and Meise call network-based APIs. These enable different systems to act as a kind of composite system. The main system architecture, in which network-based APIs are employed, are client-server architectures or

[69] Frank, Strugholtz, and Meise (2021):pp. 15, 18–19, 24; Shishmano, Popov, and Popova (2021):p. 130.
[70] Frank, Strugholtz, and Meise (2021):pp. 15, 17.
[71] Raj et al. (2020):p. 13.
[72] Perrey and Lycett (2003):pp. 116–117.

distributed applications architectures. ERP systems are a prime example of distributed systems, that are connected via network-based APIs. Analysing the system architecture and landscape is a key in deploying API based connections.[73]

One of the main benefits and use cases for APIs is a process known as API transformation. This process describes the use of APIs to assist companies whose core IT systems are lagging behind the state of the art. The goal of this transformation is to drive the development and use of APIs in the company. Thereby, the company will gain the ability of integrating current technology into an ageing systems landscape.[74] However, for this to work the APIs have to be developed with the systems landscape in mind, and resource requirements for the development are high.[75]

However, there are distinct risks associated with an API transformation and the implementation of APIs in general. These risks originate primarily from a system architecture that is not suited to handle the increased load that APIs imply in the systems. As an API of a system is implemented by another system, the load on the system that provides the API increases. When using APIs to directly integrate legacy systems into new applications, a stark increase in the workload through API calls may significantly impact the legacy system. When the business value of each API call is unclear, this poses a risk to the cost of IT operation. Furthermore, designing APIs specifically to implement legacy systems may lead to a tight coupling of the system, which ultimately increases the company's reliance on the legacy system. Generally, the coupling between two systems should be kept low, but for legacy systems this applies especially. Another downside of an integration approach where the system hosting the data is the same as the one providing the API is that the availability of the API is dependent on the system hosting the data.[76] Additionally, integrating legacy systems in this way may be highly resource intensive, when points such as low coupling between systems should be maintained.[77]

To avoid these problems, Thoo and Pezzini recommend the use of APIs provided by an integration hub.[78] This Integration Architecture was explained in section 2.1. The proposed architecture leverages APIs to integrate the back-ends with the hub and the hub to the services.[79]

A model related to this is the API economy, which defines three roles. The first role is the API provides, who provides information about their business assets. This information is then

[73] Frank, Strugholtz, and Meise (2021):pp. 19–21.
[74] Frank, Strugholtz, and Meise (2021):pp. 18–19.
[75] Thoo and Pezzini (2018):p. 6; Frank, Strugholtz, and Meise (2021):p. 19.
[76] Thoo and Pezzini (2018):pp. 3, 6.
[77] Perrey and Lycett (2003):p. 119.
[78] Thoo and Pezzini (2018):p. 7.
[79] Thoo and Pezzini (2018):p. 5.

integrated and processed by the API user. This role is also called the API consumer, as he consumes the data provided and builds his own services on these data. The last role in the model is the end user or customer.[80]

2.4.3 Major types of APIs

Three categories of APIs can be defined in an API-ecosystem. In ascending order of the degree of visibility these are private-, partner-, and public APIs.[81] **Private** APIs usually require authentication to access them as they are only designed for use inside the company that provides it.[82] The main use case is system integration. This sort of integration can lead to companies being able to build new services on legacy systems, thereby enabling new business opportunities.[83] **Partner** APIs are similar to this, as they are only intended for selected users and are not publicly disclosed but enable other businesses to access an API.[84] This selected disclosure entails administration duties, which will impact the value of the API.[85] However, new business opportunities like co-innovation, and co-development of business models and sales channels should outweigh the administration associated with managing partner APIs.[86] **Public** APIs are intended to be used by third parties, such as developers and software architects, which are probably unknown to the company.[87] Therefore, the opportunities are similar to partner APIs. However, since the public does not require the API to be free of charge, additional revenue streams can be created from companies that leverage their data as intellectual property and provide it through paid APIs. The practice of charging for API access is also used in Partner APIs, however this is done with less standardised pricing models and contracts.[88] As Palma et al. discovered, the degree of visibility affects the quality of the design of different APIs. The analysed partner and public APIs are more often designed to match patterns for APIs, while the difference in quality between these two is statistically negligible. However, the private APIs analysed in the paper tend to be qualitatively worst designed.[89]

[80] Shishmano, Popov, and Popova (2021):pp. 130–131.
[81] Frank, Strugholtz, and Meise (2021):pp. 79–80; Palma et al. (2022):p. 20; The grey literature defines three additional categories for APIs. These are open-, composite-, and unified-APIs. However, open APIs are a sub type of public APIs, whilst composite- and unified-APIs simply combine multiple APIs. (See Simpson (2022):p. 1) Therefore, this thesis will focus on the established three categories.
[82] Palma et al. (2022):p. 20.
[83] Frank, Strugholtz, and Meise (2021):p. 80.
[84] Palma et al. (2022):p. 20.
[85] Frank, Strugholtz, and Meise (2021):pp. 81–82.
[86] Frank, Strugholtz, and Meise (2021):p. 81.
[87] Frank, Strugholtz, and Meise (2021):p. 82.
[88] Frank, Strugholtz, and Meise (2021):p. 83.
[89] Palma et al. (2022):p. 25.

3 Integration Architecture: Digital Decoupling

Enterprise Integration relies on an architecture, that determines how systems interact with each other.[1] Therefore, an architecture should determine how closely or loosely systems are coupled and what the coupling depends on. This chapter will start by defining what the terms Enterprise Architecture and Integration Architecture stand for in the context of enterprise Data Integration in section 3.1. This footing is needed to analyse two Integration Architectures in sections 3.2 and 3.3. The central architecture analysed will be Accentures Digital Decoupling covered in section 3.2. This finding is achieved through a careful analysis of the patent for Digital Decoupling in the section. Section 3.3 will go into an Integration Architecture similar to Digital Decoupling, which will lead to this thesis finding similarities between the two architectures and a comparison of these architectures to models proposed in other literature.

3.1 Enterprise Architecture

> "The purpose of Enterprise Architecture is to optimise across the enterprise the often fragmented legacy of processes (both manual and automated) into an integrated environment that is responsive to change and supportive of the delivery of the business strategy."[2]

This definition of Enterprise Architecture is part of TOGAF[3]. This framework is considered to be one of the main resources for Enterprise Architecture. TOGAF covers a very wide spectrum, of which only some parts are applicable to this bachelor's thesis. However, to understand the Digital Decoupling Integration Architecture, first the phases of what TOGAF calls the ADM[4] have to be understood. The ADM is an idealised model, which may need to be tailored to the enterprise in which it is to be deployed. However, the phases of the method are still widely accepted to be at the core of Enterprise Architecture. The phase requirements engineering is the only phase that continues throughout the process described by the ADM. The other phases are executed according to fig. 3.1.

Other sources define Enterprise Architecture as a description of an enterprise.[5] This definition is relevant due to its brevity and simplicity, although it seems to neglect the effort of Enterprise

[1] The Open Group (2022):chapter 1.1.
[2] The Open Group (2022):chapter 1.1.
[3] The Open Group (2022).
[4] The Open Group (2022):chapter 1.2.2.
[5] Gulledge (2008):p. 270.

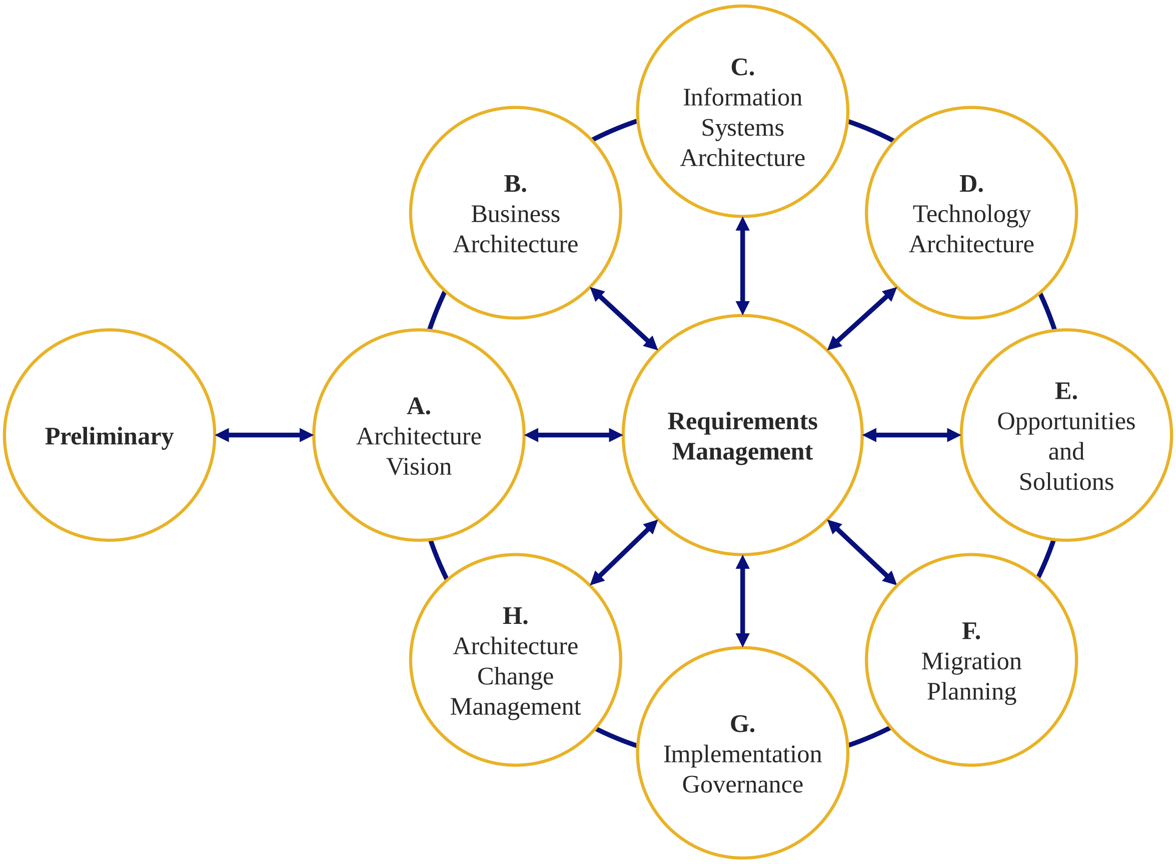


Figure 3.1: The Architecture Development Cycle *(Own representation based on The Open Group (2022):chapter 1.2.2)*

Architecture to enable the business described. The authors of this definition go on to point out, that a vision or plan of some sort is highly relevant to Enterprise Architecture.[6] This vision is derived from the business strategy through the business processes.[7] Deriving a technical System Architecture from a business strategy seems to be delusional. However, the overall strategy defines the business processes, that in turn determine the systems that need to be put in place to support or enable the business processes. In spite of this, the relationship is unidirectional, as the systems provided also affect the kind of processes that can be executed and, thereby, the business strategy that can be pursued.

A similar approach to Integration Architecture and by extension Enterprise Architecture is defined in SAP's ISA-M framework. Here, there are four levels to the architecture. The architecture is also unidirectional. Therefore, the technology or systems provided can be seen as the foundation of the business. Based on this foundation, solutions are built for distinct business areas, such as

[6]Gulledge (2008):p. 270.
[7]Gulledge (2008):p. 267.

Customer Relationship Management (CRM) systems. These solutions are then used in business processes that enable the enterprise to function.[8] Again, this definition is very closely related to SAP products and the capabilities of these products. Compared to Gulledge's definition, however, similarities in the structure arise. These similarities are especially prominent in the role that is given to integration and the effects of this integration on the business processes and by extension the business strategy.

3.2 Digital Decoupling Integration Architecture

The term Digital Decoupling is to be attributed to Accenture, as the company holds the patent[9] describing this system architecture[10]. As the term is relatively new and caters to a niche market, there is a distinct lack of literature on this topic. Aside from the patent, which describes the technical architecture of a Digital Decoupling System, no scientific papers or chapters in other specialist literature could be found in the research phase of this bachelor's thesis. Therefore, this model needs to be analysed by comparing it to different publications that either aim to solve the same problem or are concerned with similar Integration Architectures. This section will therefore lay out Accenture's idea of Digital Decoupling as described in the patent and first- and third-party web articles, whilst the following sections will compare this model to other publications.

The term Digital Decoupling is dependent on the term coupling, which is one of the core factors for measuring how well a system is designed.[11] Therefore, eliminating coupling between instances in software where possible and thereby allowing changes in one instance without needing to change other instances is one main goal of software architecture.[12] Especially in integration, decoupling is a central goal. Decoupling systems through a central system that exchanges messages in itself is no new concept in the realm of Enterprise Integration.[13] Moreover, the specialist literature even proposed ways of avoiding a central system for a hub-and-spoke approach to integration, which would be able to forgo the high complexity of a single central system.[14] The term Digital Decoupling as patented by Accenture seems to be unspecific, as it actually describes a tangible system architecture to execute this decoupling. This also entails that

[8] Allgaier (2022):p. 11.

[9] Zeier, Wagenknecht, and Kleber (2021).

[10] Accenture does not describe Digital Decoupling as an Integration Architecture, but as a System. However, as defined in section 3.1 the term architecture is more fitting. Therefore, the term Digital Decoupling System will describe the Digital Decoupling System, as shown in fig. 3.2 and the Term Digital Decoupling (integration) architecture will describe the whole approach.

[11] Fowler (2001):p. 102.

[12] Fowler (2001):p. 1.

[13] Hohpe and Woolf (2004):pp. 324–326.

[14] Hohpe and Woolf (2004):p. 325.

Accenture's proposed Integration Architecture is but one solution to a core problem of real-world Enterprise Integration.

In the words of the inventors at Accenture, "Digital Decoupling techniques [. . .] enable existing computing systems to run in parallel with new computing technologies."[15] The goal of Digital Decoupling however is not simple co-existence of these systems but integration of legacy- and new systems. The main challenge is that companies are often highly dependent on legacy systems. For some of these systems, minor updates and other revisions may be viable up to a certain point. However, these minor revisions will only support some new technologies, processes, and business strategies. Furthermore, migrating these systems to a more modern solution may be very resource and time-consuming.[16] The key concept of Digital Decoupling Systems is to convert data from existing systems into events that can be consumed by new applications. Therefore, no alterations need to be made to the existing system's data handling.[17] The result of this approach, the systems are coupled to the Digital Decoupling System instead of other systems. This leads to Integration Architecture, as described in section 2.1. In the long run, a Digital Decoupling System can serve one of two different roles according to Accenture: The first is as a new master database and the second is as an interim solution until a replacement system is stable and in production.[18]

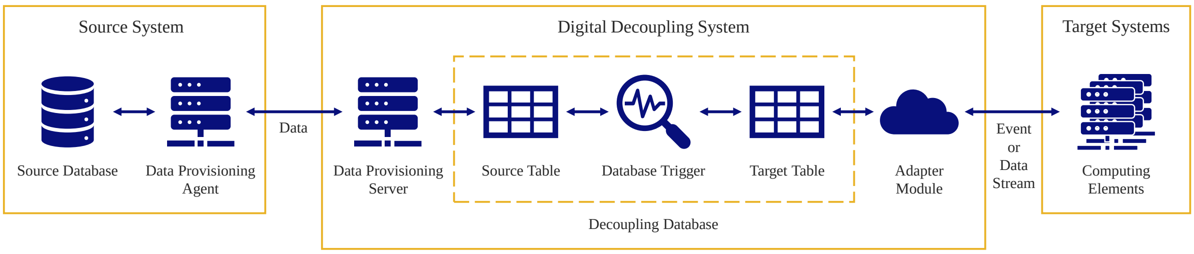


Figure 3.2: Integration Architecture of a Digital Decoupling System *(Own representation based on Zeier, Wagenknecht, and Kleber (2021):Sheet 1)*

fig. 3.2 lays out how the Digital Decoupling System is to be integrated to transform the data from the source system into messages for target systems.[19] Although there are three systems described, one of which is the Digital Decoupling System, the modules necessary for the system extend beyond this central system. The first key part of a Digital Decoupling Architecture is the data provisioning agent in the source system. In essence, this system acts as a Change Data Capture

[15] Zeier, Wagenknecht, and Kleber (2021):p. 1.
[16] Zeier, Wagenknecht, and Kleber (2021):p. 1.
[17] Zeier, Wagenknecht, and Kleber (2021):p. 2.
[18] Zeier, Wagenknecht, and Kleber (2021):p. 3.
[19] Zeier, Wagenknecht, and Kleber (2021):p. 2.

(CDC) system. The data provisioning agent is designed to push this change data from the source system into the corresponding data provisioning server when a user defined criteria for a change in the data is met.[20]

Following this, the data provisioning server writes the data to the decoupling database. More precisely, the Digital Decoupling Systems may have one or more source tables, of which the data is written to one. Therefore, the data stored in the source table will be identical to the data in the source system's database that meet the propagation criteria of the CDC system. Furthermore, the source table may document changes to the data it receives.[21] Similarly, to the data provisioning agent, the database trigger acts as a CDC system for the source table inside the decoupling database. When a change in the source table is detected, the data may be transformed and written into the target table. Transformation of the data is not certain, as the source and target formats may be identical.[22] The data written in the target table may not be persistent, as the data stored here is only used by the adapter module to create events or data streams for the target systems.[23] To create these, the data need to be transformed again from the target table schema into the schema required by the events or data streams.[24]

Aggregating the different computing systems into a group called target systems is one of the alterations to the figure described in the patent. This alteration aligns the Digital Decoupling Architecture with more standard integration systems, as described in section 2.2. However, an event may be propagated to more than one computing element.[25] The events provided by the adapter module form the *set of all events* E. A given *computing element* c may have a subset of *subscribed events* s, so that $\forall c \exists E \exists s : s \subseteq E$.[26] The concept of computing element makes the Digital Decoupling Architecture flexible, as a cloud computing element can be anything from cloud services, to on-premise systems, and even user interfaces of systems.[27]

3.3 Supporting decoupling: The Digital Integration Hub

The specific Digital Decoupling Integration Architecture described in section 3.2 is not covered by current research published in relevant journals. However, similar models that aim to achieve the same goal, have been described by other authors. This is due to Accenture's model being

[20] Zeier, Wagenknecht, and Kleber (2021):p. 2.
[21] Zeier, Wagenknecht, and Kleber (2021):p. 3.
[22] Zeier, Wagenknecht, and Kleber (2021):p. 3.
[23] Zeier, Wagenknecht, and Kleber (2021):pp. 3–4.
[24] Zeier, Wagenknecht, and Kleber (2021):p. 3.
[25] Zeier, Wagenknecht, and Kleber (2021):p. 3.
[26] Zeier, Wagenknecht, and Kleber (2021):p. 4.
[27] Zeier, Wagenknecht, and Kleber (2021):p. 3.

based on a hub-and-spoke integration model, which has been highly researched in the past.[28] Furthermore, this topic has recently gained interest and has been at the centre of a framework called the DIH architecture[29].[30] Still the topic is in its infancy, as know-how and products that would be needed to realise this architecture in enterprises are still scarce.[31]

Figure 3.3 describes the DIH architecture similar to fig. 3.2. This model was originally developed by Gartner's Thoo and Pezzini[32] but since has been used in the specialist literature[33] and real world projects[34] as a reference model for DIHs. Gartner's analysts have updated the paper outlining the model in 2020[35] and archived the original paper from 2018[36]. However, as some aspects were first mentioned by the 2018 paper and remain unchanged in the 2020 paper, this bachelor's thesis will cite the 2018 paper where no alterations in the 2020 paper could be found. Gartner refreshed both papers last in 2019 and 2022 respectively, further adding to the situation. However, these refreshes are small revisions, that do not impact the validity of the papers. Gartner's DIH model is included in SAPs ISA-M starting with version 3.3. as a cross use case integration methodology, as can be seen in fig. 2.1.[37] Therefore, the model is best described by papers provided by Gartner. However, to remain true to scientific principles, it is necessary to compare the model to other publications on the topic.

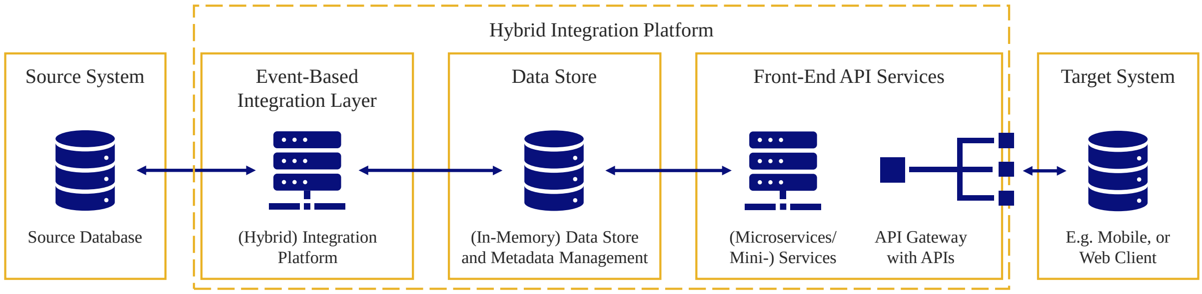


Figure 3.3: Integration Architecture using the Digital Integration Hub model *(Own representation based on Thoo and Pezzini (2018):p. 5; Pezzini and Thoo (2020):p. 3)*

Three distinct marks can be found in the DIH as laid out in fig. 3.3[38] The first is the data store in which data from the back-end systems is stored centrally. Second are the front-end API services

[28]Hohpe and Woolf (2004):pp. 324–326; Kiwon et al. (2021):pp. 54–58.
[29]Thoo and Pezzini (2018):pp. 4–5.
[30]Kiwon et al. (2021):p. 335.
[31]Pezzini and Thoo (2020):p. 7; Kiwon et al. (2021):p. 347; Allgaier (2022):p. 130.
[32]Thoo and Pezzini (2018):p. 5.
[33]Kiwon et al. (2021):p. 337.
[34]Kiwon et al. (2021):pp. 345, 347.
[35]Pezzini and Thoo (2020).
[36]Thoo and Pezzini (2018).
[37]Allgaier (2022):pp. 8, 27, 34.
[38]Kiwon et al. (2021):p. 337.

which allows the target systems to access the data in the data store. And third are the integrations of the back-end systems, which normally are based on multiple different types of integrations.[39] One omission made in fig. 3.3 is the connection of an optional analytics component[40] to the data store, as this component is not relevant for the integration scenarios covered in this thesis. This analytics component and capabilities for administration, monitoring, and security are also listed by Gartner as key building blocks for a DIH.[41] However, this distinction is not supported by other sources in specialist literature.

In a DIH architecture, the data store assumes a similar role to the Digital Decoupling System in a Digital Decoupling Architecture. It acts as a central point in the Integration Architecture that stores data from multiple source systems for low-latency access to the data without affecting the source systems.[42] Like the Digital Decoupling System, the data store therefore reduces the load on the source system and streamlines the integrations. Furthermore, in both cases, the recommended central database is an in-memory database and both systems recommend SAPs HANA database.[43] However, Gartner's model also explicitly specifies a metadata management solution within the data store, which coexists with the data store itself.[44] Overall, the technical idea based on the hub-and-spoke Integration Architecture leads to a collection of similarities between these two approaches.

The DIH architecture has advantages for three different use cases given by sources in the literature, practical guidelines, and Gartner themselves. The first is the reduction of load on the back-end systems. The reason for this reduction is that the data store acts as a kind of cache, whereby data only has to be retrieved once from the source system and can be accessed multiple times with low latency by multiple front-end applications.[45] The second advantage is a central point for API management. Managing API of distributed systems has proven to be highly complex. To combat an opaque conglomeration of APIs the DIH provides a central point for managing APIs.[46] Providing capabilities for analytics or BI is the third advantage. Since the data store of the DIH aggregates data from various production systems of an enterprise, it is a predetermined point to generate information using BI techniques.[47] A fourth point Gartner

[39] Thoo and Pezzini (2018):pp. 8–10.
[40] Thoo and Pezzini (2018):p. 5; Pezzini and Thoo (2020):p. 5.
[41] Pezzini and Thoo (2020):p. 5.
[42] Thoo and Pezzini (2018):p. 4.
[43] Thoo and Pezzini (2018):p. 6; Zeier, Wagenknecht, and Kleber (2021):p. 4.
[44] Thoo and Pezzini (2018):p. 8.
[45] Thoo and Pezzini (2018):pp. 4, 6; Kiwon et al. (2021):p. 336; Allgaier (2022):pp. 91, 130.
[46] Thoo and Pezzini (2018):p. 6; Kiwon et al. (2021):pp. 336–337; Allgaier (2022):p. 130.
[47] Thoo and Pezzini (2018):p. 11; Kiwon et al. (2021):p. 337.

continuous to give as an advantage for DIH is that the systems are designed to operate 24/7 and enable high-throughput with low-latency.[48] On the one hand, these claims are neither backed up by other sources, nor proven by any sources in the literature reviewed for this bachelor's thesis. On the other hand, proving these claims can not be done easily and definitely, as there are numerous offerings for in-memory databases available on the market. Furthermore, the actual implementation of such a system into a given IT landscape will greatly affect the performance.

The main difference between Gartner's DIH and Accenture's Digital Decoupling Architecture lies in the fact that the Digital Decoupling System transforms data from the source system into events that are consumable by the target systems. Furthermore, Accenture's patent is more concise about the technical implementation. However, the goal and the capabilities of both architectures are astonishingly similar. This goes so far that some articles by Accenture and Gartner even use very similar language. This is especially prominent when it comes to the topic of decoupling front-end systems from the data of back-end systems.[49]

Especially when reviewing the goal of decoupling front-end systems from back-end systems, the similarities between the two systems outweigh differentiating factors. The goal is very clearly formulated as enabling the modernisation of legacy systems by loosely coupling them to modern front-ends.[50] That decoupling is one of the main goals of the DIH architecture is further manifested by synonyms for this technology. The DIH is also known as the Architecture Layer, the (In-Memory) Cache Layer, or the Decoupling Layer.[51] According to Gartner, decoupling to modernise is also one of the main points that should drive the adoption rate of DIH over the coming years. This effect will be amplified by the cost savings in the source systems enabled by decoupling these systems from front-ends.[52] However, independent sources point out, that for now implementing DIH architectures is resource and know-how intensive. Furthermore, enterprises with high integration requirements currently benefit most from implementing these architectures. These high integration requirements may materialise when a growing number of accesses to the source systems is expected, or a central management for APIs is needed in an enterprise.[53]

[48] Thoo and Pezzini (2018):p. 4; Pezzini and Thoo (2020):p. 10.
[49] Thoo and Pezzini (2018):p. 4; Accenture Banking and Israel (2019).
[50] Thoo and Pezzini (2018):p. 11.
[51] Kiwon et al. (2021):p. 336.
[52] Pezzini and Thoo (2020):p. 7.
[53] Kiwon et al. (2021):pp. 343, 346.

4 Enterprise integration at Olympus

This chapter will discuss the to-be Integration Architecture at Olympus, based on the principals lied out by the specialist literature in the previous chapters. This will enable this thesis to generate a deep understanding of key drivers for the new Integration Architecture. Furthermore, it will expose key aspects where this Integration Architecture differs from the models proposed in the literature and what effect this will have. For this first, Olympus' global organisation and the resulting fragmented IT landscape will be explained in section 4.1. Following this, Olympus to-be Integration Architecture will be introduced as the Global Integration Hub in section 4.2 This will start to address some of the frame works and Integration Architectures used. With this concise overview of this architecture, section 4.3 will introduce the five key drivers for the to-be architecture, which are a result of Olympus' fragmented IT landscape on a global level and Olympus' transformation into a global, leading med-tech company. Finally, section 4.4 will derive the to-be architecture from these key drivers and critically assess this architecture to the models reviewed in chapter 3. For this, the foundational technologies used will be assessed and put into the context described in chapter 2.

4.1 Olympus' globally distributed business

Olympus is the world's leading manufacturer of optical and digital precision technology. Due to the spin-off of the scientific solution division and the carving out of the consumer business, Olympus focuses solely on products and services for the medical sector. Solutions developed by Olympus enable thousands of medical procedures, making people's lives healthier, safer, and more fulfilling.[1] The company's competencies are as diverse as its customers, each of whom has different expectations and ideas about the solutions it offers. To better meet these customer needs, Olympus is set to transform into a global, leading med-tech company. This transformation also affects the Olympus IT, where roles, teams, and systems are to be unified and integrated on a global scale. Before this transformation, the five major subsidiaries of Olympus[2] all had separate business and IT organisations. This resulted in inconsistent processes, separate teams, infrastructure, systems, and -architectures for the five major regions. Each of the five major regions are SBCs which are responsible for all tasks not related to the manufacture of Olympus

[1]Olympus Europa SE & CO. KG (2022):p. 1.

[2]Olympus' five major subsidiaries are the Sales Business Centres (SBCs) for each of the five major regions for Olympus business. These regions are in descending order by revenue: Olympus Corporation of the Americas (OCA), Olympus Europa SE & CO. KG (OEKG), Olympus Tokyo (OT), Olympus China (OCN), and Olympus Olympus Asia and Oceania (OAO). *(See Olympus Europa SE & CO. KG (2022):p. 1.)*

products. For these tasks, every major region has further subsidiaries for certain sub-regions or countries, which in some cases have their own IT teams. To a lesser extent, the fragmented IT landscape extends to the subsidiaries, as their subsidiaries and Manufacturing Business Centres (MBCs) of these regions are integrated into the ERP and BI systems of the major region to some extent.

The fragmented regional legacy IT landscapes and current lack of integration inhibits Olympus not only in delivering sound IT solutions to the customers on a global level, but also hinders the business from using the data and processes inside Olympus' systems to the extent that is required by a changing market.[3] Factors like changing expectations towards the IT capabilities of Olympus, especially regarding the capabilities to run the business globally and integrate with third parties more seamlessly, are amplifying this change.

4.2 The Olympus Global Integration Hub

Olympus' Global Integration Hub is based on different frameworks and Integration Architectures with the goal of enabling global integration, at Olympus. The integration system will be provided as an Integration Platform as a Service (iPaaS) leveraging a Digital Decoupling layer. The Digital Decoupling layer will be enabled through an API management platform provided by MuleSoft and leverage the Olympus global Data and Analytics (DnA) platform for Data Integration.

This section will analyse the resulting Integration Architecture using the frameworks established in chapter 3. For this, the components of Olympus' Global Integration Hub mentioned in the previous paragraph will be analysed and placed in the context provided by the specialist literature. One of the main frameworks used to define Olympus' Global Integration Hub architecture is the SAP ISA-M framework. The ISA-M framework as laid out in fig. 2.1 is also used to define which parts of the Global Integration Hub are in focus. Furthermore, the framework is used to outline which use case patterns and integration styles are covered by which team at Olympus, as the task of integrating systems is not solely designated to the global integration team. Especially Data Integration, which is at the centre of this thesis, is enabled by the work of the BI team on the global DnA platform.

[3]Global IT – Solution Delivery – Shared Services – Integration (2022):pp. 3–4.

4.3 The five key drivers

Five key drivers for the development of a global DIH at Olympus have been identified as part of the project brief for the according project[4].

Driver for DIH	Description
Enable the global organisation	Integration services will be provided by a central, global instance at Olympus. This will lead to a higher degree of governance.
Liberate data via APIs	APIs will provide a central point for globally distributed consumers from different backgrounds to access Olympus' data securely.
Use future-proven technology only	To enable digital transformation, only modern tools with long-term perspective will be used to build reliable integrations.
Reduce time to market	Realising integrations on a central HIP will streamline integrations, while reducing costs.
Support new business models	Reusable APIs, event-driven integration, and microservices will enable new business models that depend on integrations with Olympus entities and third-party providers.

Table 4.1: The five key drivers for the development of a global DIH at Olympus *(Own table based on Global IT – Solution Delivery – Shared Services – Integration (2022):p. 3)*

These key drivers are primarily developed out of pain points that arise with the current Integration Architecture at Olympus. Furthermore, they are based on analysis from market analysts such as Gartner and consulting agencies such as Accenture. The validity of these points therefore is easily verified when reviewing white papers provided by these companies[5]. These articles have been reviewed and placed in the context of the current specialist literature in chapter 3. However, an analysis is needed to determine whether these five key points are met with the current plans of the project.

Some of the five key drivers are validated more easily than others. For example, when examining the point use future-proven technology, only the tools to be used can be compared to proven

[4] Global IT – Solution Delivery – Shared Services – Integration (2022):p. 3.
[5] Thoo and Pezzini (2018); Pezzini and Thoo (2020); Zeier, Wagenknecht, and Kleber (2021).

frameworks such as Gartner's Magic Quadrant. In conjunction with the Olympus DnA platform, MuleSofts API management component are proposed for process integration and the Integration Architecture of the Olympus Digital Integration Hub. In Gartner Magic Quadrant, MuleSoft is one of the leaders, with the third-highest score in completeness of vision and the second-highest score in ability to execute.[6] Furthermore, MuleSoft is involved in advancing the industry by publishing papers and informational articles in their area of expertise.[7] These factors lead to a platform that has the pedigree to deliver a well-strategised long-term perspective. In conjunction with the DnA platform, which is realised using Azure Synapse Analytics, this forms a sound technological foundation which is future facing. In an analysis of Azure's cloud database management capabilities, the platform scored overall the highest in completeness of vision and was second to Amazon Web Systems offerings in ability to execute.[8] Furthermore, this system enables HIPs and class leading integration into the widely distributed Microsoft ecosystem.[9] Choosing these platforms should also affect the reliability of integrations. However, as this factor is mostly dependent on the actual implementation, verifying this claim would be beyond the scope of this thesis.

The point liberate data via APIs is based on the foundation of using future-proven technology only. This is because MuleSofts Anypoint software will be used for two distinct roles. Firstly, it will be used as a decoupling system providing modern and secure API-based access to Olympus' data assets and applications, as detailed in chapter 3. In this role, the software will act as an alternative to a harmonised global ERP system. The software offers a compelling case for this role due to the dozens of standard connectors provided by the system that simplify integration with different business applications and proprietary formats (e.g. SAP iDocs, RFCs, ...). Furthermore, Olympus' Data Lake solution based on Microsoft Azure services will provide a means of global Data Integration. This will be enabled by ETL processes on software provided by Qlik and Theobald Software. Second, both systems will provide the data through the API management service provided by MuleSoft. This builds the central API gateway from which data can be accessed by different consumers. MuleSofts API management capabilities enable this on a technical level, and the DIH-like architecture that is to be used should further promote this key driver of change. However, the actual success of introducing an API strategy depends on the actual implementations and re-use of common Data APIs by later business projects with integration demand.

[6] Pillai et al. (2021):pp. 3, 13.
[7] Pillai et al. (2021):p. 14.
[8] Cook et al. (2021):p. 3.
[9] Cook et al. (2021):pp. 19–20.

One of the business goals of realising a Global Integration Hub is to reduction in the time to market for new projects and business initiatives with related integration demand. The global alignment of integration approaches and integration governance will drive this change. For this, the fragmented back-end systems need to be decoupled from newly introduced global applications by moving and harmonising the regional data in the DnA data layer to deliver data solutions without a global ERP-harmonisation. However, the ability to reuse code has the potential to decrease the time for delivering integration solutions even more acutely by allowing the reuse of APIs and events. This will be realised by MuleSofts Anypoint Exchange Marketplace, which will provide APIs and auto-generated code stubs.

The vision is that the three points mentioned so far and the global transformation of the Olympus IT, and, by extension, the Integration team will come together to enable the global organisation. As with the time to market, this will primarily be enabled through the use of a central integration hub that can be accessed globally from any system. This push for a singular global platform is in line with the global transformation of the Olympus IT organisation and the Olympus corporation as a whole. Provided all relevant data is stored in the Global Integration Hub and the change management to transition the users of integrations to the global platform, this key driver will be fulfilled. The same goes for the goal of increasing the degree of governance over integrations. As determined in chapter 3, one of the main goals of introducing a central integration and API management platform can be establishing more control over the integrations and APIs.

However, linked to this key driver is another goal that is threatened by technologies used. To become a global leading med-tech company, Olympus set out to increase the operating profit margin to 20%. This goal should be achieved primarily by realising new business opportunities in the medtech space and reducing operating costs. This also affects the Olympus IT. Covering the full extent of these cost reduction measures on Olympus' IT is outside the scope of this thesis. However, there is a clear trade-off between platform costs and administration costs for the platforms linked to Global Integration Hub. Governing and maintaining distributed integration efforts without a central platform is resource intensive. This effect is also present at Olympus.[10] However, to realise this resource saving, that could turn into cost savings, a new platform needs to be introduced. This entails one-time costs for introducing these platforms and ongoing costs for licencing and maintaining the platforms. This is particularly relevant, as both main platforms used for Olympus' Global Integration Hub are found to have significant licencing costs and opaque licencing models compared to their competition.[11]

[10] Global IT – Solution Delivery – Shared Services – Integration (2022):p. 4.
[11] Pillai et al. (2021):p. 14; Cook et al. (2021):p. 20.

Assessing the ability of a Global Integration Hub at Olympus to support new business models is difficult. This is mainly due to the uncertainty that is associated with new business models.[12] However, following the industry trend of API driven and event-based integration should lead Olympus to be able to leverage best practises and integration patterns pioneered by other enterprises. Furthermore, the integration of third-party services, which is a clear goal of this key driver, should be streamlined by following these industry standards. The ambiguity of this key driver for a Global Integration Hub is one of the reasons use cases for new Data Integrations need to be defined and analysed. This will enable an assessment of the integration demand at Olympus in the near future.

4.4 The to be architecture

The to be Integration Architecture at Olympus employs different architectural concepts and frameworks for data- and process integration. This section will clarify how these concepts work in conjunction to cocreate what Olympus calls the Global Integration Hub[13], as displayed in fig. 4.1.

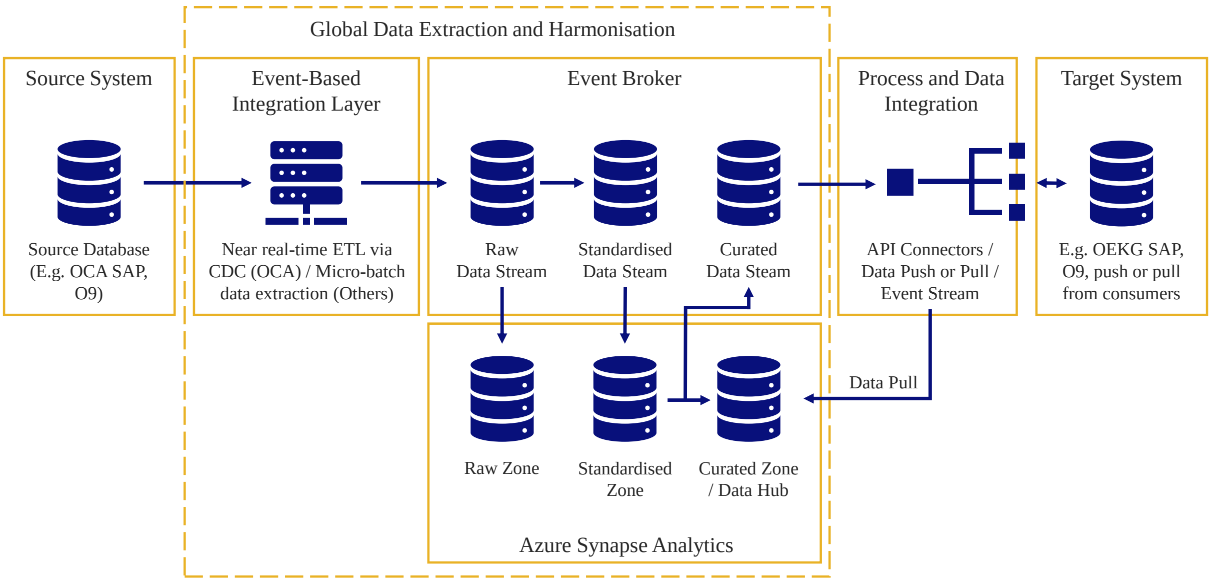


Figure 4.1: Olympus' to be Integration Architecture utilising the Global Integration Hub *(Own representation based on Global IT – Solution Delivery – Shared Services – Integration (2022):p. 10; Tanck (2022):p. 1)*

As the name suggests and fig. 4.1 clearly shows, with the central event broker, Olympus to be Integration Architecture is based on a hub-and-spoke integration model. There are two

[12] Rajnoha et al. (2014):p. 166.
[13] Global IT – Solution Delivery – Shared Services – Integration (2022).

distinct reasons that make this integration pattern particularly viable. First, the global integration team seeks to cut down on costs for managing heterogeneous integrations in each region. Section 2.1 found the effect of a reduced number of required connections using a hub-and-spoke style integrations. A reduction in connections reduces the required management, provided the complexity of integrations is not altered. Second, defining having a common data model for integrations enables cross-regional integrations with reasonable effort. The curated data stream in the event broker and the curated zone on the Azure system can therefore act as the global common ground for enterprise data. This enables streamlined integrations with reusable code for APIs and reusable data. In this role, the Global Integration Hub almost takes on the role of a global master data system or a globally harmonised ERP system that can be used by all regions. This characteristic makes it particularly interesting for business initiatives introducing new applications on a global level at Olympus instead of using direct connections to all regions and their specific ERP systems. Having a major enterprise system deployed globally from the start would be a first for Olympus.

However, fig. 4.1 does not show that the event broker is segmented into regions and a global master system.[14] This approach is common in integration systems based on a central event broker or hub and is called a “Message Broker hierarchy”[15]. Splitting up systems prevents the central event broker from becoming too complex and, therefore, unmaintainable. For decoupling systems, this approach offers the added advantage of decoupling the subnets.[16] In the case of Olympus, this allows the company to maintain the low degree of coupling of the systems across the regions. In the interest of decoupling systems, this is the right step, but it makes a global harmonisation of systems and processes more difficult. Here, Olympus’ goals of becoming a globally integrated leading med-tech company collide with the reality of the regional systems landscapes developed historically.

Another technology employed by the Global Integration Hub is the data lake in form of the raw zone on the DnA platform. This part can be seen as part of an analytics or BI system coupled to this architecture. This architectural decision is similar to Garters DIH but lacks a high performance in-memory data store and the integration of an optional analytics component. However, in the case of the Olympus global integration platform, the globalDnA platform is used to enable Data Integration to utilise the work of the BI team, which has already integrated many Olympus data sources on a global scale for reporting purposes.

[14] Tanck, Korniyasov, et al. (2022):p. 1.
[15] Hohpe and Woolf (2004):p.325.
[16] Hohpe and Woolf (2004):p. 325.

As part of the DnA implementation project, the ERP systems of OEKG, OT, OCN, and OAO were connected using micro-batch data extraction. OCA's ERP platform was connected using a near-real-time ETL or CDC engine. The O9 solutions Supply Chain Management (SCM) system will probably be connected using an API based ETL process. This special position of O9 solutions is because the platform will be used as a pilot for Data Integration using the Global Integration Hub.[17] Therefore, the End-to-End (E2E) SCM project, which aims to optimise global SCM processes by introducing O9 as the new SCM platform for Olympus globally, is a major stakeholder of the Global Integration Hub. As a global platform and a major strategic initiative, the E2E SCM project also takes on a kind of lighthouse position for future projects and Olympus' system landscape going forward. This guiding position can also be seen in the process and Data Integration of the system.

The part of the architecture called process and Data Integration is considerably simplified in fig. 4.1. Therefore, fig. 4.2 gives a detailed view of how the target systems are integrated. Figure 4.2 shows that the main way to propagate the data will be by outgoing communication from the curated data stream derived from the DnA data lake. The preferred approach that will be pioneered in the E2E SCM project is that the data be propagated as event streams to which the target systems can subscribe, as described in section 3.2. As a fallback for systems that do not support event based integration, a data push approach will be implemented in MuleSoft by mapping the events to target system APIs. Further APIs will be provided to these systems for API-led back-end integration. The ERP systems of each region are integrated through a process API and preferably a system API or connectors. Both of which will be implemented in MuleSoft. This integration will be the part of the system that enables the integration hub to act as a central point for all regions to connect to to integrate their data with one another globally. For this, the System based on MuleSoft will also be able to look up reference data from the standardised zone of the DnA data lake. At first, this integration will be used to integrate the SCM systems reporting into the regional ERP systems. This integration permits the write back of data from the global SCM system to the local ERP systems. However, once this architecture is established, and an proven, further integrations of this kind are set to proceed the E2E SCM project. The other means of data propagation is by target systems pulling data from the curated zone through MuleSofts API Gateway and data APIs. This approach should enable web and mobile applications that integrate Olympus' enterprise data. However, access latencies also have

[17] Global IT – Solution Delivery – Shared Services – Integration (2022):p. 7.

to be kept low enough for these applications. The current technology deployed at Olympus does not meet these latency requirements.[18]

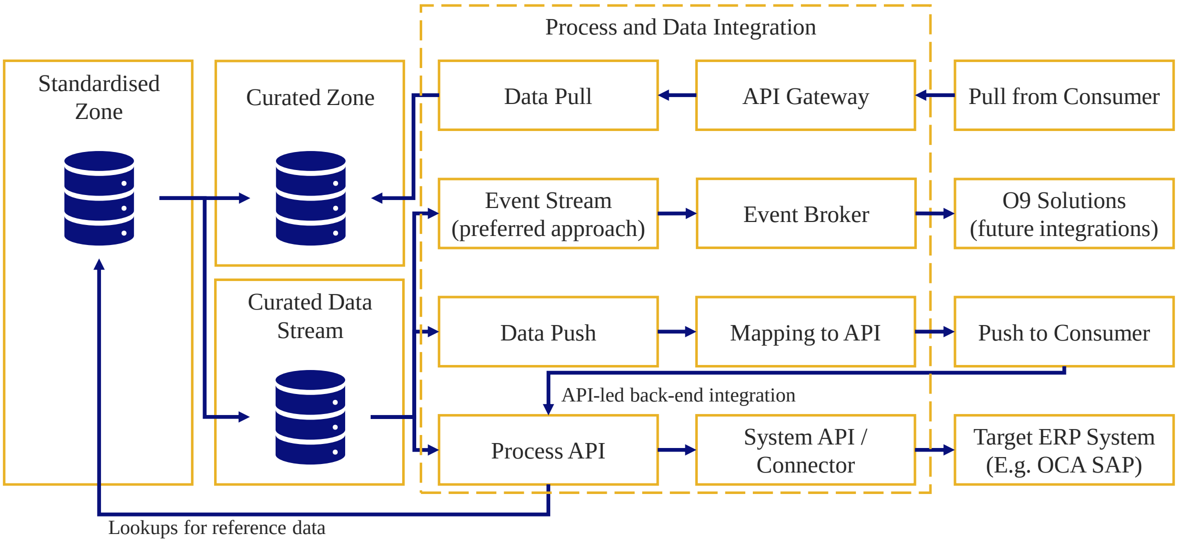


Figure 4.2: Process and Data Integration from the integration hub to target systems *(Own representation based on Tanck (2022):p. 1)*

Although the Olympus Integration team set out to design the global Integration Architecture according to the DIH Integration Architecture and Accentures Digital Decoupling, the actual decoupling of systems is not explicitly described in the internal documentation. This is even though the points which serve the decoupling in both of the aforementioned Integration Architectures are realised in the Olympus Integration Architecture. Namely, these points are the event-based integration layer and the process and Data Integration, respectively. This similarity is especially clear from the unified presentation in figs. 3.2, 3.3 and 4.1. Even the source tables as described by Accenture for a Digital Decoupling System are present as the raw data zone in the Olympus Global Integration Hub. Also, target tables as described by Accenture are implemented. However, the target table is split into the Curated Data Stream, which only stores data for seven days, and the curated zone, which could store data indefinitely, provided enough storage capacity. The possibility of storing data in the target tables only for a short amount of time is also described by Accenture. However, here the storage duration is not fixed. Rather data is deleted, after it is sent to the target system.

The point at which Gartner and Accenture propose the connection of source systems to the integration system is realised as the event-based integration layer laid out in fig. 4.1. This point is equivalent to the event-based integration layer Gartner proposes in fig. 3.3 and roughly equivalent

[18] Habermann (2022).

to the data provisioning server Accenture uses in fig. 3.2. The DIH architecture proposes a purely event-based integration. However, realising this at Olympus involves CDC engines coupled to the source system, which act like the data provisioning agents in a Digital Decoupling Integration Architecture except that they send events and not data.

The point at which the target systems are integrated with the Global Integration Hub is designed similarly to a DIH. Especially, bidirectional communication between the final databases and target system along with the use of APIs and event streams makes this connection clear. Furthermore, the front-end API services are enabled through the use of a gateway for API management. These APIs may also be provided to partners and publicly after aligning with critical stakeholder like the Olympus IT security team and the Olympus IT compliance team.[19] Bidirectional communication through events in the Digital Decoupling Integration Architecture. However, the communication between the target system and the hub is only realised through events. This contrasts Olympus' approach, where nearly every system that does not pull data and is not O9 solutions will be integrated via data push through API. This is expected to change over time, as the integration approach pioneered in the E2E SCM project is rolled out to other projects. However, over the coming years, this difference will persist.

The point at which the source and target systems are actually decoupled can be described as the central data store in each Integration Architecture. Figure 4.1 shows, that this point is in the transformation between the raw data stream and the standardised data stream. In a DIH this point is in the data store and Accentures Digital Decoupling Systems defines the decoupling database as the point at which the decoupling occurs. When reviewing all three systems it is clear that Olympus approach of using the Data Integration system provided by DnA leads to a more convoluted system. However, since the system is proven and already deployed globally, this provides a good starting point for an Integration Architecture. Eliminating a tight coupling within the integration system and a relatively tight coupling of the source and target systems to the integration system is the same approach as pursued by the DIH and the Digital Decoupling Architecture. As discussed in chapter 3, this allows back-end and front-end systems to be maintained, altered, and even replaced with less effort than direct integration of the system.

[19] Tanck, Palaniappan, and Korniyasov (2022):p. 1.

5 Defining use cases

This chapter is focused on defining clear use cases for Data Integration in a Digital Decoupling Integration Architecture. For this, the knowledge gained about Integration Architectures and the Global Integration Hub at Olympus as laid out in chapters 3 and 4 respectively is used. This knowledge is used to set the boundaries of the use cases and contribute to the chosen research methods.

5.1 Research Method

The research method used to define use cases will be based on two pillars:

1. Document reviews of key materials of the projects that provide the use cases and
2. Semi-structured expert interviews with key stakeholders of the projects that are familiar in the technical realisation and the business case or business value of the given project.

The document review is a critical first step in developing an appropriate structure for each interview. This enables the interviews to focus on the knowledge of experts that is not captured by the documents. Therefore, no information that can be easily obtained by other means will be produced in these interviews.[1] The definition of what is considered to be a document is somewhat vague, as many kinds of documents can be reviewed.[2] However, the documents reviewed in this thesis are mainly project sources outlining the project goal and some technical boundaries. Therefore, the review of the documents will extract relevant information on the business requirements of the projects in terms of Data Integration and the system used. This information can be retrieved well from these internal documents.[3]

The expert-interview was chosen for this thesis, as it supplements the more objective nature of the literature review and document review with subjective and contextual information that is specific to the situation at Olympus. This is possible through the tactic knowledge that the experts have about the company structure and application landscape at Olympus.[4] Nevertheless, giving the expert interview semi-structured boundaries allows for the necessary objectivity and succinctness of questions to use this research method in this thesis. However, even with this

[1] Baur and Blasius (2019):p. 682.
[2] Sharma and K. Pandey (2013):p. 37.
[3] Sharma and K. Pandey (2013):p. 37.
[4] Baur and Blasius (2019):p. 681.

measure and the expert's special role, the knowledge obtained is still not objective and has to be understood correctly, as it cannot be measured unambiguously.[5]

As Baur and Blasius points out, the situation in which the interviews are conducted and, by extension, the roles given to the interviewer and the interviewee are critical to obtaining a tangible result from this qualitative research method.[6] The role of the expert should be considered very carefully, as the expert is to be chosen based on relevant knowledge to the question.[7] For the interviewer, this mainly implies giving the interviewee the maximum amount of freedom to answer the questions in their own words and to be open to the views of the interviewee.[8]

To keep the conditions in the interviews uniform and to create an undisturbed and familiar setting, the interviews conducted for this thesis are carried out as online meetings in the first language of the interviewee. Adjusting the structure of the interview according to the results of the document review should lead the interview situation to focus more on one topic, creating a constant environment, contributing to the value of the information gathered.

The appendix includes a transcript of every interview. Additionally, an introduction of the expert is also included, together with a justification of the expert status given to the person.

5.2 Evaluation of results

The possible initiatives elected to define the use cases are expected to benefit from Data Integration in a Digital Decoupling Integration Architecture at Olympus. This expectation is derived from the claims made in the specialist literature reviewed in chapter 3. Here, web applications and other applications that aim to integrate modern front-ends are the main beneficiary of this Integration Architecture. Another use case is based on the E2E SCM project that has been introduced as a path breaking project for this Integration Architecture. However, this project also offers an interesting alternative use case, as the system which is integrated is not described by the specialist literature as a system that can clearly leverage the advantages of this Integration Architecture.

Each of the following subsections presents the results of the document review and expert interview and is named after the main use case for Data Integration derived from these results. A tabular overview of each use case is provided in appendix A.

[5] Baur and Blasius (2019):pp. 682–683.
[6] Baur and Blasius (2019):pp. 669, 682.
[7] Baur and Blasius (2019):p. 671.
[8] Baur and Blasius (2019):pp. 672, 682.

5.2.1 Digital Customer Experience

On a high level, the goal of the Go to Market (GTM) project is to optimise the Digital Customer Experience (DCX).[9] To achieve this, a unified DCX will be created globally to offer unified digital services and an unique user experience to Olympus customers via different channels.[10] The main services that fall under this DCX are in the fields of pre-, and post-sales, BI, as well as a knowledge database for the customers, and order tracking for newly ordered products and repairs.[11] The order tracking for repairs is especially important for Olympus, as the business model is increasingly dependent on full service contracts. As these contracts include regular maintenance and check-ups of the products, a transparent customer experience is essential to meet customer expectations and advance Olympus as a business. This customer experience is already being delivered by other companies in different industries, as these services are becoming more common and therefore become an expectation of the customers.[12] However, this experience is not required to have real time Data Integration, as long as the data provided to the customer by the DCX is as timely, as the information he can receive through other means, like contacting a sales representative.[13]

For this tracking of repairs and newly ordered products, data from core Olympus systems must be integrated into the front-end system the GTM project is set out to create. In this case, the core system is the back-end system in the Integration Architecture laid out in section 4.2 and the DCX platform the front-end, or user-facing system. As the relevant data for order tracking, pre-, and post-sales is captured in the Sales Force instances, these CRM systems are the back-end systems. Additionally, data is stored in the ERP systems, which also makes these systems relevant as back-end systems. When reviewing the front- and back-end systems, the fragmented landscape of the Olympus IT and business structure becomes apparent. There are 17 Sales Force instances and numerous front-end systems that could leverage data from these instances.[14] The front-end systems are so highly fragmented and for the most part to be replaced by the GTM, that no clear number was determined for the project. However, the number of applications is expected to be in the hundreds, with over 200 websites and microsites.[15] The possible front-end systems extend beyond the needs of the GTM project, as will be demonstrated in more use cases.

[9] Digital CX Global Team (2021):p. 3.
[10] Digital CX Global Team (2021):pp. 12–18.
[11] Digital CX Global Team (2021):p. 18.
[12] Digital CX Global Team (2021):p.7; Sheridan (2022).
[13] Sheridan (2022).
[14] Digital CX Global Team (2021):p. 8.
[15] Digital CX Global Team (2021):p. 8; Sheridan (2022).

As the DCX platform is set out to create a globally unified platform, data from all regions and therefore all production Sales Force instances needs to be integrated. Furthermore, the Data Integration needs to be performant and independent of the maintenance intervals of the back-end systems and adaptable to new innovation.[16] All of these aspects have been identified as advantages of Integration Architectures based on the DIH or Digital Decoupling Architecture in sections 3.2 and 3.3. As the GTM project expects the success of the DCX to be highly dependent on integration with this sort of back-end system, it would highly benefit from this global, decoupled Integration Architecture, especially for creating a globally unified DCX.[17]

5.2.2 Web and mobile applications

Analysing documents for a concrete use case for Data Integration for web and mobile applications at Olympus proves difficult. This is due to two reasons:

1. There is currently no project in this domain which needs to integrate data from distributed systems. Even though there are future plans for projects in this domain, no extensive documentation or prototyping has been done.

2. The integration solutions previously used by the Marketing Solutions teams, which is responsible for web and mobile applications, were custom build by the team for extracting and in some cases caching data from ERP and CRM systems.[18]

However, the Marketing Solutions team is building the Infection Prevention website[19], which is set to be the precursor to new web developments at Olympus. Reviewing a document that describes the technology stack used shows the integration of Olympus systems such as the Media Asset Management system.[20]

Furthermore, this use case of Data Integration for the Infection Prevention site is particularly interesting, as this site's architecture is well suited to the Olympus Global Integration Hub. This is because the site is designed as a static page, which reduces the majority of the API calls to the build time of the website. In the current version of the website nearly all API calls are executed on-build with the search function on the website being the only API call which is executed at the run-time.[21] This leads to the timeliness of data being of a lesser importance. Furthermore,

[16] Digital CX Global Team (2021):p. 6.
[17] Sheridan (2022).
[18] Schrader (2022).
[19] Olympus (2022).
[20] Schrader et al. (2022):p. 1.
[21] Schrader (2022).

it reduces the latency requirements for the integration platform. However, future iterations of the website may call APIs during the run-time.[22] This would necessitate a highly performant integration platform like the Global Integration Hub to deliver a suitable experience for websites.

A media asset management system and a potential Product Information Management system are the main systems of interest for the integration of these web services.[23] Therefore, the first versions of the Global Integration Hub will not address the main Data Integration use case of these solutions. However, the interview also revealed that integrations with the systems in scope for the Global Integration Hub are also highly relevant for web and mobile applications. The team responsible for creating these applications developed and maintained an integration and caching service specifically because of the execution times in the realm of multiple minutes for Data Integrations with the ERP systems.[24]

5.2.3 Global Supply Chain Management

The ultimate business goal of the E2E SCM project is to introduce global solutions for integrated business planning and a control tower at Olympus.[25] It is one of the major strategic initiatives which aim to transform Olympus into a global, leading med-tech company. Section 4.4 covers the implications of integrating this system on a global level with the ERP systems at Olympus.

The use cases for Data Integration in a Digital Decoupling Integration Architecture are also not explicitly described in the literature provided by Gartner or Accenture. However, this use case is highly important for Olympus, as this Integration Architecture is planned to be rolled out to other projects, after a successful first attempt with the E2E SCM project.[26]

The integrations needed to deliver an integrated business planning solution globally based on the fragmented system landscape are a central element to the success of the project. To accomplish the tasks of an E2E SCM tool, integrations with the core ERP systems of each region are the very nature of this project. Of the 89 source systems for relevant data that were identified, over 42% are the various SAP ERP central components, which are a subset of the 74 of core applications needed.[27] These numbers clarify the need for a central integration platform like a Digital Decoupling System, as designing and managing these without a central integration platform would have been uneconomical.

[22] Schrader (2022).
[23] Schrader (2022).
[24] Schrader (2022).
[25] Michel et al. (2022):p. 2.
[26] Arai and Umezawa (2022):p. 1.
[27] Morohashi, Albrecht, and Michel (2022):pp. 3–4.

The most relevant finding from the interview with Achim Habermann[28] was that the source systems, from which data is being integrated into O9, are constantly changing and with relatively high frequency. This change was expected from the document review and can be attributed to the project being in early phases. However, the frequency with which the systems relevant to the business change is high enough that during the initial phases of the E2E SCM project, there were multiple changes in what systems are relevant beyond the scoping phase.[29] Furthermore, the large number of systems in OT and OEKG alone adds to the complexity of creating a sound global platform for Data Integration.

Another key factor of this project is the requirement for a globally harmonised common data model, because all regional source ERP applications need to feed their data into the standard data model of one global SCM system. This data model builds the foundation of the integration hub and will be reused by future projects. Designing this data model as part of the integration platform and based on the data model provided by DnA marks the importance of this project for the Global Integration Hub. However, it may also affect the data model in a way, that future projects can only obtain limited functionality from the Global Integration Hub.[30]

5.3 Discussion

A central point all use cases have in common is that the interviews confirmed that they have experienced, and/or are expecting the back-end systems providing the data to change. Furthermore, every expert interviewed attributed this change to the fragmented global system landscape at Olympus.[31] Both of these points were also discovered in the document reviews. Therefore, the ability to decouple the systems of these applications from the back-end systems is almost required to meet the needs of these use cases. However, only in the case of the E2E SCM project was the impact of the fragmentation on the system clear and quantifiable. In this case, the document review uncovered the impact of global Data Integration on core business systems, which confirms the role of the project as a demonstration of this Integration Architecture for future projects. Having a central global integration platform, which is able to decouple the fragmented back-end systems from global systems which may be customer facing, therefore offers great business value to Olympus. This was especially true for use cases regarding web and mobile applications and the GTM project. Both of these initiatives expressed the need for

[28] Habermann (2022).
[29] Habermann (2022).
[30] Habermann (2022).
[31] Habermann (2022); Schrader (2022); Sheridan (2022).

"pain of glass"[32] or a "Fassade"[33] as a sort of abstraction layer for integrations. This would be delivered by the Global Integration Hub.

Customer facing solutions, which defined the use cases for Data Integration into web systems, are highly reliant on short response times with varying patterns to data consumption. The Digital Decoupling Integration Architecture used is more performant in these regards than the back-end systems it relies on. However, as the analysis of the Olympus Global Integration Hub in section 4.2 revealed and the interview with Achim Habermann confirmed, designing the Global Integration Hub on top of the DnA system may have a negative impact on the hub in the future.[34] Using the global harmonised data model on the DnA system can lead to negative repercussions in the long term. This is due to the DnA system and, by extension, the timeliness of the data in it being designed for BI tasks. Therefore, data may be outdated by minutes. Even with good access latency, this makes the system undesirable as a singular solution for some customer facing systems and web applications.[35] For these use cases the main benefit would be accessing data from enterprise back-end systems such as the central ERP or CRM systems with latencies in hundreds of milliseconds and not minutes.[36] The GTM project, had the same need for low access latencies and query times. However, in this case, the timeliness of the data was not as important as the consistency of the timeliness across all channels to the customer.[37]

However, future developments on the DnA system are bound to introduce a near real time ETL process, which would greatly increase the data timeliness. Depending on the definition of near real time, this would eliminate most of the problems concerning data timeliness. Furthermore, designing the Global Integration Hub as a system built on the data provided by the DnA system has the upside of gaining value from the work done for this system for a second time. Furthermore, the systems fit well into the plans proposed for decoupling and the idea of a Digital Decoupling System. However, the implications of using a BI systems as the basis for a Global Integration Hub ultimately leads the DnA system to perform tasks it was not designed to perform. Decoupling back-end systems from front-end services is a strong point, but meeting the different integration needs of the front-end data consumers like web applications proves to be difficult. Therefore, only some of the advantages of a Digital Decoupling Integration Architecture or a DIH are leveraged in this form.

[32] Sheridan (2022).
[33] Ger. facade. See Schrader (2022).
[34] Habermann (2022).
[35] Schrader (2022).
[36] Schrader (2022); Sheridan (2022).
[37] Sheridan (2022).

The research carried out in the course of this thesis was carefully executed according to proven methods. However, by reviewing how the results presented in the previous sections may not reflect the actual situation, the following threats to the validity of the results must be acknowledged.

A key bias that applies to the research carried out in this thesis is the confirmation bias. The projects for the use cases were primarily selected based on cases that the specialist literature predicted to take advantage of this change in the integration technology. However, an attempt to cancel out this bias is the use case based on the E2E SCM project, as cases like these are not mentioned in the literature reviewed in this thesis. For the scope and goal of this thesis, this counterpoint alongside the careful review of the use cases offers a reasonable defence against this bias.

Another threat to the validity of the research done in this thesis may be that only qualitative research methods have been used. However, this fact is based on the nature of this research topic. As defining use cases for Data Integration in a Digital Decoupling Architecture at Olympus has to be based on the projects actually aiming to use these integrations, quantifiable data would have to be obtained from the qualitative basis of these ongoing projects. In cases where quantifiable data was able to be obtained and contribute to this thesis, it was used to assist the qualitative research.

6 Conclusion

This thesis analysed the impact of the Digital Decoupling Integration Architecture implemented by the Global Integration Hub at Olympus. For this, three distinct use cases for Data Integration were defined. An analysis of these use cases has shown in which areas the use cases can already be covered by the implementation of the Integration Architecture. Furthermore, shortcomings due to the data timeliness of the Global Integration Hub have been found.

In concussion, the document reviews and expert interviews conducted for this thesis were used to define three use cases for Data Integration in a Digital Decoupling Integration Architecture at Olympus. These three use cases are the development of a DCX platform for the GTM project, enabling web and mobile applications to have to integrate Olympus' enterprise data, and decoupling the new global SCM system from the fragmented IT landscape at Olympus. All these use cases were found to benefit from the main advantages of Digital Decoupling. In descending order of the business value for these use cases advantages include: Decoupling global systems from the fragmented and localised IT landscapes, developing a single point of truth for enterprise data across regions with near real time data timeliness, decreasing the time to market for integrations by accelerating data ingestion and processing times, and delivering a flexible integration concept that is highly compatible. The corresponding business value of these advantages were also derived from the document reviews and expert interviews.

Designing the Olympus Global Integration Hub to make use of the DnA system further increased the business value of this Integration Architecture by leveraging work already done to harmonise data globally. However, it also forms the basis for the weaknesses of this system. Due to the current state of the DnA system, real time Data Integrations with this system are impossible. Olympus needs to accelerate the development this system to meet business expectations towards the Global Integration Hub. This recommendation leads to answering the research questions posed in the introduction of this thesis:

RQ 1: By allowing Olympus to introduce new global applications decoupled from the fragmented IT landscape caused by the former regional approach, the Digital Decoupling Integration Architecture supports the efforts to unify Data Integration globally. This thesis presented different integration technologies and integration approaches, which make this unification possible. Most offer additional business value, because the unification also leverages economies of scale inside Olympus and reduces operating- and maintenance costs. A prime example of this is the global consolidation of integration tools and integration experts. Another example is the global consolidation of ETL tools, the DnA system, and the curated zone of this system. As this consolidation goes as far as consolidating tools and expertise between the BI and Integration teams, this is a particularly strong case for integration technologies supporting Olympus' efforts towards globally unified Data Integration. Using these systems to decouple the global target systems from the fragmented source systems, is the first principal of enabling unified Data Integration for Olympus globally.

RQ 2: Three use cases for Data Integration in the Global Integration Hub were defined. These use cases are not a holistic list of all possible use cases, but rather a focused selection of specific areas in which this Integration Architecture offers benefits compared to other approaches. All use cases represent global initiatives at Olympus, that need to integrate enterprise data. The first use case defined is the development of the DCX platform as part of the GTM project. This platform is mainly concerned with integrating data from the 17 Sales Force instances Olympus operates globally. Using Digital Decoupling techniques for this project is advantageous, as this data will be used in future integrations, and as the project team wants to use an abstraction layer for Data Integration. The second use case is based on the Infection Prevention website as a lighthouse project for future web developments at Olympus. The Expert Interview for this use case discovered that the team developing these solutions designed a sort of Data Integration and Caching system similar to the Global Integration Hub. This demonstrates the need for this kind of Integration Hub that decouples front-end systems from back-end systems. However, developing this Integration Hub once at Olympus would prove to be more economical than designing integration systems for each project. Furthermore, the Infection Prevention website would also benefit from being able to easily integrate more data into the website, such as ERP data or a future product information system. Using the Global Integration Hub for this use case would allow for this. The third use case is the biggest beneficiary of the Global Integration Hub. It was concerned with introducing a global E2E SCM platform based on O9 solutions software. 89 ERP- and CRM system need to be integrated with this system. As these systems are integrated frequently into other enterprise systems of Olympus, using the Global Integration Hub as an integration platform, is expected to accelerate future projects integrating with these systems. Furthermore, the ERP- and CRM systems deployed at Olympus are bound to change over the coming years, as replacement projects for the ERP- and CRM systems of certain regions are underway. Using an Integration Architecture designed with loose coupling as a central goal will help with these transitions.

The use cases defined in this thesis gave a clear understanding of the capabilities of the Global Integration Hub at Olympus and, by extension, Digital Decoupling Integration Architectures. Even with the shortcomings of utilising the current version of the DnA platform as a layer for Data Integration, the system still has many valid use cases. Olympus should drive more projects to use the Global Integration Hub to leverage the reuse-effects of this platform. The use cases analysed in this thesis show a clear benefit for projects utilising this Integration Architecture. Further research on the topics of Integration Architectures should lead to findings regarding the value of Digital Decoupling to businesses. With the world becoming more data driven, integrating data will be one of the key challenges in enabling the modern world. The distinct lack of recent independent research published in relevant journals about Data Integration should be addressed. Major trends like Big Data, Thing Integration for Internet of Things devices, and a trend towards automation seem to lead to more research in this field. However, topics such as Integration Architectures are mainly developed by analysts and for-profit companies instead of being part of the scientific discourse. Additionally, an in-depth analysis of the quantifiable impact of utilising the Global Integration Hub as the integration platform for a single project should yield a profound understanding of the use of this platform's business value.

Literature

Abelló, Alberto (2015). "Big Data Design". In: *Proceedings of the ACM Eighteenth International Workshop on Data Warehousing and OLAP*. DOLAP '15. Association for Computing Machinery: New York, NY, USA, pp. 35–38. ISBN: 978-1-4503-3785-4. DOI: 10.1145/2811222.2811235. URL: https://doi.org/10.1145/2811222.2811235 (visited on 10/30/2022).

Allgaier, Matthias (2022). *SAP Integration Solution Advisory Methodology (ISA-M) Version 4.0*. Version 4.0. SAP SE: Walldorf. 150 pp.

Alserafi, Ayman et al. (Oct. 4, 2017). "DS-Prox: Dataset Proximity Mining for Governing the Data Lake". In: *Lecture Notes in Computer Science* (October 2017), pp. 1–10. ISSN: 978-3-319-68473-4. DOI: 10.1007/978-3-319-68474-1_20. URL: https://www.researchgate.net/publication/320074046 (visited on 10/26/2022).

Baur, Nina and Jörg Blasius (2019). *Handbuch Methoden der Empirischen Sozialforschung*. Springer Vieweg. in Springer Fachmedien Wiesbaden GmbH: Wiesbaden, GERMANY. ISBN: 978-3-658-21308-4. URL: http://ebookcentral.proquest.com/lib/nordakademie/detail.action?docID=5716629 (visited on 10/22/2022).

Chang, She-I (Jan. 2004). "ERP life cycle implementation, management and support: Implications for practice and research". In: *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*. 37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the, pp. 1–10. DOI: 10.1109/HICSS.2004.1265556.

Chaudhuri, Surajit and Umeshwar Dayal (Mar. 1, 1997). "An overview of data warehousing and OLAP technology". In: *ACM SIGMOD Record* 26(1), pp. 65–74. ISSN: 0163-5808. DOI: 10.1145/248603.248616. URL: https://doi.org/10.1145/248603.248616 (visited on 10/27/2022).

Cook, Henry et al. (Dec. 14, 2021). "Magic Quadrant for Cloud Database Management Systems". In: *Magic Quadrant*, pp. 1–48. URL: https://www.gartner.com/en/documents/4009286 (visited on 11/18/2022).

Cuzzocrea, Alfredo and Paolo Ciancarini (Aug. 13, 2021). "SeDaSOMA: A Framework for Supporting Serendipitous, Data-As-A-Service-Oriented, Open Big Data Management and Analytics". In: *2021 5th International Conference on Cloud and Big Data Computing (ICCBDC)*. ICCBDC 2021. Association for Computing Machinery: New York, NY, USA, pp. 1–7. ISBN: 978-1-4503-9040-8. DOI: 10.1145/3481646.3481647. URL: https://doi.org/10.1145/3481646.3481647 (visited on 10/17/2022).

Fang, Huang (June 2015). “Managing data lakes in big data era: What’s a data lake and why has it became popular in data management ecosystem”. In: *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*. 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), pp. 820–824. DOI: 10.1109/CYBER.2015.7288049.

Fowler, M. (July 2001). “Reducing coupling”. In: *IEEE Software* 18(4). Conference Name: IEEE Software, pp. 102–104. ISSN: 1937-4194. DOI: 10.1109/MS.2001.936226.

Frank, Roland, Sebastian Strugholtz, and Fabian Meise (2021). *Bausteine der digitalen Transformation: Wie APIs Unternehmen den Weg in die Programmable Economy ebnen*. Springer Fachmedien Wiesbaden: Wiesbaden. ISBN: 978-3-658-35106-9. DOI: 10.1007/978-3-658-35107-6. URL: https://link.springer.com/10.1007/978-3-658-35107-6 (visited on 11/02/2022).

Funk, Burkhardt et al. (2010). *Geschäftsprozessintegration mit SAP*. 1. Auflage. Springer Berlin Heidelberg: Berlin, Heidelberg. ISBN: 978-3-642-12720-5. DOI: 10.1007/978-3-642-12721-2. URL: http://link.springer.com/10.1007/978-3-642-12721-2 (visited on 10/25/2022).

Groombridge, David (2022). *Top Strategic Technology Trends 2023*. 2023rd ed. Gartner: Stamford. 33 pp. URL: https://emtemp.gcom.cloud/ngw/globalassets/en/publications/documents/2023-gartner-top-strategic-technology-trends-ebook.pdf (visited on 12/12/2022).

Gulledge, Thomas R. (2008). “Architecture-driven Enterprise Integration”. In: *International Journal of Management and Enterprise Development* 5(3), p. 265. ISSN: 1468-4330, 1741-8127. DOI: 10.1504/IJMED.2008.017433. URL: http://www.inderscience.com/link.php?id=17433 (visited on 10/18/2022).

Hohpe, Gregor and Bobby Woolf (2004). *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions*. Addison-Wesley Professional: Boston. 683 pp. ISBN: 978-0-321-20068-6.

Kabiri, Ahmed and Chiadmi Dalila (July 1, 2013). “Survey on ETL processes”. In: *Journal of Theoretical and Applied Information Technology* Vol. 53 (July 2013). ISSN: 1817-3195. URL: https://www.researchgate.net/publication/256761671 (visited on 10/28/2022).

Kadadi, Anirudh et al. (Oct. 2014). “Challenges of data integration and interoperability in big data”. In: *2014 IEEE International Conference on Big Data (Big Data)*. 2014 IEEE International Conference on Big Data (Big Data), pp. 38–40. DOI: 10.1109/BigData.2014.7004486.

Kimball, Ralph and Joe Caserta (2004). *The data warehouse ETL toolkit: practical techniques for extracting, cleaning, conforming, and delivering data*. Wiley: Indianapolis. 491 pp. ISBN: 978-0-7645-6757-5. URL: https://ia800206.us.archive.org/15/items/2004TheDataWarehouseETLToolkitRalphKimball/2004%20-%20The%20Data%20Warehouse%20ETL%20Toolkit%20%28Ralph%20Kimball%29.pdf (visited on 10/28/2022).

Kiwon, Adam et al. (2021). *SAP-Schnittstellenmanagement: der praktische Leitfaden für die hybride Integration*. 1. Auflage. SAP PRESS. Rheinwerk Verlag: Bonn. 512 pp. ISBN: 978-3-8362-7982-6.

Knoblock, Craig A and Pedro Szekely (2013). "Semantics for Big Data Integration and Analysis". In: *University of Southern California*, pp. 28–31. URL: https://usc-isi-i2.github.io/papers/knoblock13-sbd.pdf (visited on 10/25/2022).

Koufteros, Xenophon, Mark Vonderembse, and Jayanth Jayaram (Jan. 13, 2005). "Internal and External Integration for Product Development: The Contingency Effect of Uncertainty, Equivocality, and Platform Strategy". In: *Decision Sciences* 36, pp. 97–133. DOI: 10.1111/j.1540-5915.2005.00067.x.

Lenzerini, Maurizio (June 3, 2002). "Data integration: a theoretical perspective". In: *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. PODS '02. Association for Computing Machinery: New York, NY, USA, pp. 233–246. ISBN: 978-1-58113-507-7. DOI: 10.1145/543613.543644. (Visited on 10/25/2022).

Mertens, Peter (2013). *Integrierte Informationsverarbeitung 1 – Operative Systeme in der Industrie*. Springer Fachmedien: Wiesbaden. ISBN: 978-3-8349-4394-1. DOI: 10.1007/978-3-8349-4395-8.

Palma, Francis et al. (July 2022). "Investigating the Linguistic Design Quality of Public, Partner, and Private REST APIs". In: *2022 IEEE International Conference on Services Computing (SCC)*. 2022 IEEE International Conference on Services Computing (SCC). ISSN: 2474-2473, pp. 20–30. DOI: 10.1109/SCC55611.2022.00017.

Perrey, R. and M. Lycett (Jan. 2003). "Service-oriented architecture". In: *2003 Symposium on Applications and the Internet Workshops, 2003. Proceedings.* 2003 Symposium on Applications and the Internet Workshops, 2003. Proceedings. Pp. 116–119. DOI: 10.1109/SAINTW.2003.1210138.

Pezzini, Massimo (Oct. 15, 2018). "Innovation Insight for Hybrid Integration Platforms". In: *Innovation Insight*, pp. 1–20. URL: https://www.gartner.com/en/documents/3891574 (visited on 11/09/2022).

Pezzini, Massimo and Eric Thoo (Oct. 7, 2020). “Innovation Insight: Turbocharge Your API Platform With a Digital Integration Hub”. In: *Foundational*, pp. 1–23. URL: https://www.gartner.com/en/documents/3987367 (visited on 10/11/2022).

Pillai, Shameen et al. (Oct. 28, 2021). “Magic Quadrant for Full Life Cycle API Management”. In: *Magic Quadrant*, pp. 1–45. URL: https://www.gartner.com/en/documents/4006268 (visited on 11/15/2022).

Plattner, Hasso (June 29, 2009). “A common database approach for OLTP and OLAP using an in-memory column database”. In: *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*. SIGMOD ’09. Association for Computing Machinery: New York, NY, USA, pp. 1–7. ISBN: 978-1-60558-551-2. DOI: 10.1145/1559845.1559846. URL: https://doi.org/10.1145/1559845.1559846 (visited on 10/26/2022).

Prema, A. and A. Pethalakshmi (Feb. 2013). “Novel approach in ETL”. In: *2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering*. 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, pp. 429–434. DOI: 10.1109/ICPRIME.2013.6496515.

Raj, Aiswarya et al. (Aug. 2020). “Modelling Data Pipelines”. In: *2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. 2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pp. 13–20. DOI: 10.1109/SEAA51224.2020.00014.

Rajnoha, Rastislav et al. (Jan. 8, 2014). “Business Information Systems: Research Study and Methodological Proposals for ERP Implementation Process Improvement”. In: *Procedia - Social and Behavioral Sciences*. 2nd World Conference on Business, Economics and Management 109, pp. 165–170. ISSN: 1877-0428. DOI: 10.1016/j.sbspro.2013.12.438. URL: https://www.sciencedirect.com/science/article/pii/S1877042813050702 (visited on 10/18/2022).

Sharma, Shreta and S. K. Pandey (Aug. 23, 2013). “Revisiting Requirements Elicitation Techniques”. In: *International Journal of Computer Applications* 75(12), pp. 35–39. ISSN: 0975-8887. DOI: 10.5120/13166-0889. URL: http://research.ijcaonline.org/volume75/number12/pxc3890889.pdf (visited on 10/28/2022).

Shaul, Levi and Doron Tauber (Aug. 30, 2013). “Critical success factors in enterprise resource planning systems: Review of the last decade”. In: *ACM Computing Surveys* 45(4), 55:1–55:39. ISSN: 0360-0300. DOI: 10.1145/2501654.2501669.

Shishmano, Krasimir Todorov, Veselin Dimitrov Popov, and Petya Emilova Popova (Oct. 2021). “API Strategy for Enterprise Digital Ecosystem”. In: *2021 IEEE 8th International Conference* *on Problems of Infocommunications, Science and Technology (PIC S&T)*. 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), pp. 129–134. DOI: 10.1109/PICST54195.2021.9772206.

Sivabalan, S and R I Minu (Nov. 2021). "Heterogeneous Data Integration with ELT and Analytical MPP Database for Data Analysis Application". In: *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)*. 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), pp. 1–5. DOI: 10.1109/i-PACT52855.2021.9696841.

Sorokine, Alexandre et al. (Oct. 31, 2016). "Big data as a service from an urban information system". In: *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*. BigSpatial '16. Association for Computing Machinery: New York, NY, USA, pp. 34–41. ISBN: 978-1-4503-4581-1. DOI: 10.1145/3006386.3006391. URL: https://doi.org/10.1145/3006386.3006391 (visited on 10/17/2022).

Sreemathy, J. et al. (Mar. 2020). "Data Integration in ETL Using TALEND". In: *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). ISSN: 2575-7288. Coimbatore, pp. 1444–1448. ISBN: 978-1-72815-197-7. DOI: 10.1109/ICACCS48705.2020.9074186.

Thoo, Eric and Ted Friedman (Dec. 22, 2017). "Infuse Your Data Hub Strategy With Data and Application Integration". In: *Foundational*, pp. 1–10. URL: https://www.gartner.com/en/documents/3840568 (visited on 10/25/2022).

Thoo, Eric and Massimo Pezzini (June 26, 2018). "Innovation Insight: The Digital Integration Hub Turbocharges Your API Strategy". In: *Foundational*, pp. 1–19. URL: https://www.gartner.com/en/documents/3880263 (visited on 10/25/2022).

Zagan, Elisabeta and Mirela Danubianu (June 2021). "Cloud DATA LAKE: The new trend of data storage". In: *2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), pp. 1–4. DOI: 10.1109/HORA52670.2021.9461293.

Zeier, Alexander, Dominik Wagenknecht, and Alexander Kleber (Feb. 18, 2021). "Digital decoupling". U.S. pat. 20210049186A1. Accenture Global Solutions Ltd. URL: https://patents.google.com/patent/US20210049186A1/en (visited on 11/06/2022).

Zhang, Hao et al. (July 2015). "In-Memory Big Data Management and Processing: A Survey". In: *IEEE Transactions on Knowledge and Data Engineering* 27(7). Conference Name: IEEE

Transactions on Knowledge and Data Engineering, pp. 1920–1948. ISSN: 1558-2191. DOI: 10.1109/TKDE.2015.2427795.

Internal Sources

Arai, Takahiro and Arisa Umezawa (June 2022). “GAEA: Globally Unified HR System”. Tokyo. URL: https://olympusssa.sharepoint.com/sites/ProjectCenter20/SitePages/GAEA.aspx (visited on 12/02/2022).

Digital CX Global Team (Apr. 15, 2021). “GTM Digital Customer Experience – DCX Journey Vision”.

Global IT – Solution Delivery – Shared Services – Integration (June 22, 2022). “Global Integration Hub, Project Brief, Implementation of a global Integration Platform as a Service (iPaaS) at Olympus”. Hamburg, Center Valley, Singapore.

Michel, Manuel et al. (Dec. 3, 2022). “Project Structure | Integrated Setup for SCM Technology Roadmap”. Hamburg, Tokyo. URL: https://olympusssa.sharepoint.com/:p:/r/sites/SCMITRoadmap/Shared%20Documents/General/10%20-%20Project%20Management/20%20-%20Project%20Structure/20201203_IT%20SCM%20Roadmap_PROJECT%20STRUCTURE.pptx?d=w75b03afa94424ec08483ac742732e069&csf=1&web=1&e=RkgcmT (visited on 02/12/2022).

Morohashi, Kiwako, Christoph-Robin Albrecht, and Manuel Michel (2022). “Attachment G Functional and Technical Scope Objects Final”. Tokyo, Hamburg, Center Valley.

Olympus Europa SE & CO. KG (Mar. 2022). “Corporate Fact Sheet EMEA (English)”. Hamburg.

Schrader, Marc et al. (Mar. 22, 2022). “Infection Prevention – Technology Stack”. Hamburg.

Tanck, Stefan (Oct. 16, 2022). “DnA Integration Architecture – Global IT | Integration – Confluence”. Hamburg. URL: https://olympus-eu.atlassian.net/wiki/spaces/INTEGRATIO/pages/2706342006/DnA+Integration+Architecture (visited on 11/22/2022).

Tanck, Stefan, Alexander Korniyasov, et al. (July 11, 2022). “Deployment Architecture v1.0”. Hamburg, Center Valley, Singapore. URL: https://olympusssaemea.sharepoint.com/:p:/r/sites/GlobalIntegrationStrategy/_layouts/15/Doc.aspx?sourcedoc=%7B6EFEFFB1-1C1C-4A08-A7C5-163A36FCF124%7D&file=Deployment%20Architecture%20v1.0.pptx&action=edit&mobileredirect=true (visited on 11/23/2022).

Tanck, Stefan, Vigneswaran Palaniappan, and Alexander Korniyasov (Sept. 19, 2022). “WP2: API Management – Global IT | Integration – Confluence”. Hamburg. URL: https://olympus-eu.atlassian.net/wiki/spaces/INTEGRATIO/pages/2681470991/WP2+API+Management (visited on 11/23/2022).

Internet Sources

Accenture Banking and Andre Israel (Mar. 27, 2019). *Digital decoupling: A maturing concept in digital disruption*. Accenture Banking Blog. In collab. with Bertt Goode, Jason Mark, and Edwin VanderOuderaa. URL: https://bankingblog.accenture.com/digital-decoupling-maturing-concept-digital-disruption (visited on 11/06/2022).

Amazon Web Systems (2022). *What is a data lake?* Amazon Web Services, Inc. URL: https://aws.amazon.com/big-data/datalakes-and-analytics/what-is-a-data-lake/ (visited on 10/30/2022).

Association for Computing Machinery (Dec. 12, 2022). *ACM Digital Library*. ACM Digital Library. URL: https://dl.acm.org/ (visited on 12/12/2022).

Google (2022). *What is a Data Lake?* Google Cloud. URL: https://cloud.google.com/learn/what-is-a-data-lake (visited on 10/30/2022).

Institute of Electrical and Electronics Engineers (Dec. 12, 2022). *Browse Popular – IEEE Xplore*. Browse Popular – IEEE Xplore. URL: https://ieeexplore.ieee.org/popular/all (visited on 12/12/2022).

Micrisoft (2022). *Data Lake | Microsoft Azure*. URL: https://azure.microsoft.com/en-us/solutions/data-lake/ (visited on 10/30/2022).

Olympus (2022). *Infection Prevention in Endoscopy – Expert Know-How*. URL: https://infectionprevention.olympus.com/en-us/ (visited on 12/07/2022).

Scott, Geoff (Jan. 20, 2021). *ASUG Pulse of the SAP Customer 2021 Results*. ASUG. URL: https://www.asug.com/insights/asug-pulse-of-the-sap-customer-2021-results (visited on 10/27/2022).

Simpson, J. (Mar. 15, 2022). *6 Types of APIs: Open, Public, Partner, Private, Composite, Unified | Nordic APIs |*. Nordic APIs. URL: https://nordicapis.com/6-types-of-apis-open-public-partner-private-composite-unified/ (visited on 11/07/2022).

The Open Group (Apr. 2022). *The TOGAF® Standard – Digital Edition, 10th Edition*. URL: https://pubs.opengroup.org/togaf-standard/ (visited on 11/04/2022).

Interviews

Habermann, Achim (Dec. 2, 2022). *Interview with Achim Habermann, Global Data Program Manager – Advanced Expert, on 02.12.2022*.

Schrader, Marc (Dec. 5, 2022). *Interview with Marc Schrader, Delivery Lead (Marketing), on 05.12.2022*.

Sheridan, Dennis (Dec. 5, 2022). *Interview with Dennis Sheridan, Americas + CX Sub-Domain Lead, on 05.12.2022*.

Satutory declaration

I, Simon Reinersmann, born on November 9th 2000, declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Hamburg, December 23rd 2022: ______________________________

Signature

Attachments

To this there thesis is a DVD attached. It contains this thesis as printed. Additionally, a version of this thesis that is laid out for viewing on computers is provided. The latter version also includes all the appendices, as they have not been printed for environmental reasons. Furthermore, the sources used in this thesis are included on the DVD. These documents are named in the following scheme '*<Last name of the first author><Publication year>*'. If the document was in a different format than PDF, it was converted to a PDF. The original format of converted files is available upon the examiners' request. Further references are available to the examiners upon request. Internal documents may be provided with the approval of the company supervisor.

The following list gives an overview of how the documents are structured on the DVD. For files in the folder *20_sources*\ only example documents are listed to convey the structure as concisely as possible.

```
.\
├─ 10_BT\
│   ├─ 10_BT print version.pdf
│   └─ 20_BT digital version.pdf
└─ 20_sources\
    ├─ 10_literature sources\
    │   └─ doe2020.pdf
    ├─ 20_internal sources\
    │   └─ doe2021.pdf
    └─ 30_internet sources\
        └─ doe2022.pdf
```

Appendices

A Overview of use cases

This chapter of the appendix serves to give an overview of the use cases defined in this bachelor's thesis. For this, the use cases are laid out in tabular form that is identical between the use cases to assist in comparing the use cases against one another.

A.1 Digital Customer Experience

Initiative	Developing a unified global DCX for the GTM project.
Business goal	The business goal of the GTM project as a whole, is to design a multichannel customer experience for medical professionals who are Olympus' customers. For this business goal, a globally harmonised DCX is to be created and implemented in all regions. The final goal of this initiative is to have a singular online presence for Olympus customers offering Olympus' services.
Target / acting system	The DCX platform developed by the GTM project.
Main source systems to be integrated	• Sales Force instances of the regions • Vertex tax engine (External) • Worldpay B2B Payments (External)

Data Integration requirements	• Low request latency (maximum in the hundreds of milliseconds) • Medium data timeliness (Integrated data may be outdated by hours or days, as long as the timeliness is constant across all channels to the customer.) • Ability to integrate data from multiple, changing back-end systems.
Advantages of realising this use case using the Global Integration Hub	The ability to unify the 17 Sales Force instances into a single point for Data Integration into the DCX is a clear advantage of a Digital Decoupling Integration Architecture. Furthermore, the requirements for low request latency and a medium timeliness of the data are almost the perfect use case for the Global Integration Hub.
Disadvantages of realising this use case using the Global Integration Hub	A disadvantage of realising this project using the Global Integration Hub could be that the project is already prepared for integrating with multiple Sales Force instances. However, integrating the 17 Sales Force instances once with the Global Integration Hub and reusing these integrations on future projects will prove more economical for Olympus as a company in the long term.
Verdict on the use case	The use case of realising the DCX using the Global Integration Hub's Data Integration capabilities built on DnA fits very well with the capabilities of this integration system. The requirements posed by the project to meet business needs are all met by the Global Integration Hub. Furthermore, the downside of altering the project plans for integrations should positively impact the GTM project and future project.

Table A.1: Use case: Digital Customer Experience *(Own table)*

A.2 Web and mobile applications

Initiative	Developing the Infection Prevention website.
Business goal	The business goal of the Infection Prevention website is to educate medical professionals on the sterilisation and best practices for handling of endoscopy products. The website serves as a knowledge base from hosting scientific articles, educational material, manuals for Olympus products, and practical guidelines for Infection Prevention. The goal of the development of the website is to pioneer a new approach to developing and operating websites at Olympus. In this lighthouse position, the Infection Prevention sites serves as a demonstration and blueprint for future web and mobile applications.
Target / acting system	The Infection Prevention website.
Main back-end system to be integrated	• Media Asset Management System • Potential future Product Information Management System
Data Integration requirements	• Medium data timeliness (Almost all data is integrated through API calls, which are executed at build-time of the static websites. Therefore, the data timeliness at build time is important.) • Low request latencies for some API calls. As the search function is live and partially accesses API data concerning Olympus products needs to be available with latencies in the hundreds of milliseconds. • A central point from which product information and media asses can be integrated into the website.

Advantages of realising this use case using the Global Integration Hub	The main advantages of realising the Data Integration of the Infection Prevention website on the Global Integration Hub are threefold: 1. The DevOps pipeline of the Infection Prevention website is very compatible with the Integration Architecture of the Global Integration Hub. Therefore, integrating the source systems into the Global Integration Hub would yield value for this website and future project accessing these systems. Additionally, the integration can rely almost exclusively on the Data Integration capabilities of the Global Integration Hub. This will keep integration efforts for this website low, whilst giving the ability to access data from other source systems like ERP or CRM systems. Having a reliable and easy way of integrating this data further strengthen the lighthouse position of this approach to developing web applications at Olympus. 2. The lighthouse function of this website for future web and mobile applications of Olympus is highly relevant for the overarching acceptance of the Global Integration Hub. Furthermore, future developments are to be based on this web architecture. Therefore, the load on the source systems of the data will increase considerably as time goes on. The Global Integration Hub is able to absorb this influx in requests. 3. Olympus aims to unify its business and, by extension, the customer facing systems globally. Utilising the Global Integration Hub as a single point for integrations can support this goal for websites built on the blueprint of the Infection Prevention site.
Disadvantages of realising this use case using the Global Integration Hub	The main disadvantage of realising the Data Integration on the Global Integration Hub is that this integration is already built on another system. This disadvantage is very prominent for the project goal of delivering the Infection Prevention website, as no benefit would be gained by transitioning this website to use the Global Integration Hub.

Verdict on the use case	Utilising the Global Integration Hub in this use case does not yield an immediate benefit for the Infection Prevention website. However, for future website built on the blueprint of the website utilising the Global Integration Hub would be highly beneficial. This is mostly due to Olympus' commitment to a global organisation and the DevOps pipeline of the website.

Table A.2: Use case: Web and mobile applications *(Own table)*

A.3 Global Supply Chain Management

Initiative	Global E2E SCM project
Business goal	The business goal of the global E2E SCM project is to introduce systems for integrated business planning and a supply chain control tower at Olympus. These systems are set out to be global systems, meaning that there will be no dedicated instances of these systems for Olympus' regions.
Target / acting system	The O9 system used for integrated business planning and the supply chain control tower.
Main back-end system to be integrated	• ERP systems of the regions • In later revisions CRM systems of the regions
Data Integration requirements	• Highly available Data Integration platform with 24/5 support. • Medium data timeliness (Preferably in the minutes to seconds). • Providing data as a global single point of truth.

Advantages of realising this use case using the Global Integration Hub	As the first system at Olympus that is designed from the ground up as a global system, O9 has a special need for the Global Integration Hub. To integrate data from the ERP- and CRM systems of the regions unidirectionally, an integration hub aggregating this data globally is almost unavoidable. Especially with changes in the ERP systems and changing data structures of the ERP- and CRM systems, having a decoupling system like the Global Integration Hub is highly advantageous for successful operation of the E2E SCM system. This is especially important as replacements for some of the ERP- and CRM systems are already being initiated.
Disadvantages of realising this use case using the Global Integration Hub	The main downside of utilising the Global Integration Hub and the DnA system for Data Integration is, that the system is trying to replace that data storage component of a global ERP system and thereby create a global data schema at Olympus. From a purely technical point of view, creating a globally unified ERP system and Master Data Management (MDM) system at Olympus would be the best solution. This would eliminate the shortcomings of realising Data Integration on a platform designed for BI tasks. However, when raising this point it should be noted, that the main shortcomings of the architecture like a data timeliness and uptimes of the system are being improved and should not affect the O9 System once it is live.
Verdict on the use case	The global E2E SCM project is highly dependent on reliable Data Integrations. As the first truly global core system of Olympus, decoupling this system from the fragmented IT landscape is important for the mid- to long term operation of the system. The Digital Decoupling techniques used by the Global Integration Hub to allow for this loose coupling of the O9 System and ERP-and CRM systems are therefore highly advantageous. Furthermore, the project integrates core business data into the Global Integration Hub. These integrations will be reused, further increasing the business value of the global E2E SCM project and the Global Integration Hub.

Table A.3: Use case: Global Supply Chain Management *(Own table)*