MESG

MESTRADO EM ENGENHARIA DE SERVIÇOS E GESTÃO

# Service Design for Data Lineage in Financial Services

*Mafalda Sofia de Almeida Gomes*

**Master Thesis**

Supervisor at FEUP: Prof. João Alberto Vieira Campos Pereira Claro

Supervisor at Euronext Technologies: Eng. Marco Machado

U.PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

2020-09-25

*To my mother, father, sister, friends, and boyfriend.*

## Abstract

Data Lineage describes the sources of a data set for products or processes, and the transformations of data from their original to their current form. It can help an organization communicate Data Quality (suitability, reliability, accuracy, redundancy) and prevents misinterpretation or misuse of data, among other benefits. The possibility of framing Data Lineage as a service suggests the interest of looking at it from a Service Design perspective. Service Design involves converting a service brief (customer needs) into a specification for both the service and its delivery and control, while reflecting the organization's options (i.e., its aims, policies, and costs). This can be done using several frameworks, methodologies, or techniques, proposed by a diverse literature.

The main goal of this work is indeed to bring together the two areas and contribute to a better understanding of how Data Lineage can be designed as a service within a company. To achieve this, a Service Design methodology, Multilevel Service Design, was used to design Data Lineage for the case of a company in the Financial Sector, taking into consideration the different levels presented in that methodology: customer experience, service concept, service system and service blueprint. Data Collection techniques were also used, not only to understand the customer experience, but also to include the main users and stakeholders of Data Lineage in the study and to obtain valuable data that was used later in the study.

After the design of the service, it was also necessary to create a plan of implementation that fitted the goals and needs of the organization, which included several presentations to the different departments of the organization, training sessions and the preparation of several documents, including an official Procedure for the creation, access, use and analysis of Data Lineage.

The use of a Service Design methodology to design and implement Data Lineage as an internal service brought added value to the organization due to its ability to analyze in depth the full capabilities of Data Lineage and allowing for a larger role to be played by its users and stakeholders.

## Resumo

O *Data Lineage* descreve as fontes de um conjunto de dados de produtos ou de processos e as transformações dos dados desde o seu estado original até ao atual. Pode ajudar as empresas a comunicar a Qualidade dos Dados (adequação, fiabilidade, exatidão e redundância) e evita a má interpretação ou utilização indevida dos dados, entre outros benefícios. A possibilidade de perspetivar o *Data Lineage* como um serviço sugere o interesse de olhar para ele de um ponto de vista de *Service Design*. O *Service Design* envolve a conversão de um resumo do serviço (as necessidades dos clientes) em especificações para o serviço e para a sua entrega e controlo, refletindo simultaneamente as opções da organização (ou seja, os seus objetivos, políticas e custos). Isto pode ser realizado através do uso de *frameworks*, metodologias ou técnicas, propostas por uma literatura diversa.

O principal objetivo deste trabalho é de facto juntar as duas áreas e contribuir para uma melhor compreensão de como o *Data Lineage* pode ser concebido como um serviço dentro de uma empresa. Para tal, foi usada uma metodologia de *Service Design*, o *Multilevel Service Design*, para desenhar o *Data Lineage* para o caso de uma empresa do Setor Financeiro, tendo em consideração os diferentes níveis apresentados nessa metodologia: experiência do cliente (*customer experience*), conceito do serviço (*service concept*), sistema do serviço (*service system*) e mapa do serviço (*service Blueprint*). Foram também usadas técnicas de recolha de dados, não só para compreender a experiência do cliente, mas também para incluir no estudo os principais utilizadores e interessados do *Data Lineage* e recolher dados valiosos também usados no estudo.

Após o desenho do serviço, foi também necessário criar um plano de implementação que fosse ao encontro dos objetivos e necessidades da organização, que inclui um conjunto de apresentações aos diferentes departamentos, sessões de formação, e a preparação de vários documentos, incluindo um Procedimento oficial para a criação, acesso, uso e análise do *Data Lineage*.

O uso de uma metodologia de *Service Design* para desenhar e implementar o *Data Lineage* como um serviço interno foi uma mais valia devido à sua capacidade de fazer uma análise aprofundada de todas as capacidades do *Data Lineage* e permitir que os seus utilizadores e interessados assumam um papel mais importante.

**Acknowledgments**

First of all, I would like to thank the University of Porto, especially FEUP (Faculdade de Engenharia da Universidade do Porto) for giving me the opportunity to spend the past five years studying with great professors who gave me so much knowledge.

I want to thank Professor João Claro for always being available and for encouraging me throughout these months, giving me motivational speeches and advices, even at times when I would struggle the most.

At the company where this work was accomplished, Euronext Technologies, I would like to thank Marco Machado and Ricardo Oliveira for always being open, ready and available to help me with all my questions and struggles, and for keeping pushing me forward and encouraging me through all this process.

I would also like to thank my friends that directly or indirectly have helped me through the years, to Francisca Campos and Inês Barbieri who since high-school have helped me and have given me some of the best days in my life, to Tiago Sequeira and Catarina Gomes for all the moments in these past five years and for all the laughs, tears and talks that we shared. And especially to Rita Carvalho, one of the best persons that I have met, for all the songs danced, for all the laughs and for all the advice and all the support in the past months.

To my family, my mother Amélia Gomes, my father Joaquim Gomes and my sister Raquel Gomes, for helping me in every bad situation, for all the car trips and sung songs, for every laugh and for making me strong and being able to live every day knowing that I will always have my family by my side.

Lastly, I need to thank my boyfriend, Tomás Carneiro. Without him, I would not be the person that I am today. He fills me with love, confidence, happiness, and strength every single day, even with just a hug. The past six years have been the best years of my life just by having him by my side, and no words can describe what he means to me.

## Table of Contents

## List of Tables

## List of Figures

**List of abbreviations**

| | |
|---|---|
| ABS | Application and Business Services |
| API | Application Programming Interface |
| BIA | Business Impact Analysis |
| CMCS | Multi-scale Chemical Sciences |
| CVC | Customer Value Constellation |
| DB | Database |
| DBMS | Database Management System |
| DMV | Department of Motor Vehicles |
| ESSW | Earth System Science Workbench |
| ETL | Extract, Transform, Load |
| FMEA | Failure Mode and Effect Analysis |
| FP | Foundational Premise |
| FTU | Facilities, Transformation, Usage |
| FX | Foreign Exchange |
| GDPR | General Data Protection Regulation |
| GOOSE | Geographic Object-Oriented Support Environment |
| IMF | International Monetary Fund |
| ISO | International Organization for Standardization |
| IT | Information Technology |
| MSD | Multilevel Service Design |
| MTFC | Multidimensional Treatment Foster Care |
| POC | Proof of Concept |
| QFD | Quality Function Deployment |
| REST | Representational State Transfer |
| SEB | Service Experience Blueprint |
| SEE | Service Encounter Experience |
| SIC | Stages of Implementation Completion |
| SQL | Structured Query Language |
| SSA | Service System Architecture |
| SSME | Service Science, Management and Engineering |
| SSN | Service System Navigation |
| TREC | Transparent Result Caching |

| | |
|---|---|
| TRIZ | Theory of Inventive Problem Solving |
| UI | User Interface |
| VCE | Value Constellation Experience |

# 1 Introduction

## 1.1 Background

This study has its original motivation on Euronext Technologies, based in Porto, more specifically on its Data Office. Inside the Data Office, the Data Governance department has the capabilities of metadata dandling and management, data sources and master data management, compliance checking and both regulatory and legislative impact matters, including matters relating, but not limited, to data privacy (more specially, data masking), data governance and data management best practices. Data Governance plays a very important role inside Euronext, since it works with very useful tools that can improve and ease the use of data and assures compliance by the organization.

IT Governance in a company refers to what decisions must be made to ensure effective management and use of IT (decision domains) and who makes the decisions (locus of accountability). On the other hand, Data Governance refers to who holds the decision rights and is accountable for an organization's decision-making about its data assets (Khatri and Brown, 2010). Data Governance may also be defined as the process of managing the availability, usability, integrity and security of the data in enterprise systems, based on internal data standards and policies that also control data usage (Search Data Management, 2020).

## 1.2 Problem Description

The main goal of this study is to understand how a framework like Data Lineage can be designed and implemented as a Service on a company from the Financial sector. Data Lineage, being "the sources and derivatives of a data set of product or (…) all the processes and transformations of data from original measurements to current form" (Bose, 2002), has the main goal inside a company such as Euronext to assist the teams when working on existing or new projects of software development, in helping them understand the type of data that they are working with and how the changes that they perform can affect the data sets downstream and inter dependent applications, assisting them also in improving their relationship with the clients. Since Euronext is a company from the Financial sector, Data Lineage can also be a powerful tool when working with reports for regulatory purposes and GDPR, by making it possible to understand where the data is and how it flows.

Before working on the design of the service with Multilevel Service Design, an expansion of the scope of Data Lineage was made. This means that, using the steps explained in the Data Lineage mappings sub-section, inside the Problem Characterization section, several mappings were created in order to have a cohesive work before designing the service.

The overall design of the new service was then done with the use of Multilevel Service Design. That way, it was possible to understand the position of the main stakeholders and how the new service would be used. Using Multilevel Service Design, the process of implementation is facilitated, thanks to its ability to design from scratch the whole service, explaining the benefits from the beginning.

In order to be able to completely achieve the goals of the company, after the design of the service using Multilevel Service Design, there is a need to implement the service, based on the

conceptual model of Aarons, Hurlburt and Horwitz (2011): Exploration, Adoption Decision/Preparation, Active Implementation, Sustainment. Each of these steps include sub-steps, based on the needs of the company.

## 1.3 Research Questions

To guide this study, ensuring its focus and purpose throughout all the work to be carried out, it is defined the following **Action Research** questions (McCombes, 2019):

- How can Data Lineage be implemented as an Internal Service, in order to mitigate regulatory risks and requirements?
- Based on the needs of a company in the Financial Sector, is Multilevel Service Design an appropriate service design methodology for such service?
- What is the best way to implement such new service?

## 1.4 Study Development at Euronext Technologies

This study has its origin in the Data Governance team, inside Euronext Technologies, based in Porto. Euronext Technologies is one of the Technology Centers of Euronext. Euronext is the leading pan-European exchange, that has a global reach and local presence, covering Belgium, France, Ireland, the Netherlands, Norway, Portugal, and the UK. It is divided in three components: **sales offices** on three continents, from which international investors and trading members can take benefit, the **regulated markets** in Amsterdam, Brussels, Dublin, Lisbon and Paris and their **technology centres**, where Euronext develop market leading solutions to enable trading in an evolving and highly demanding landscape.

## 1.5 Research Methodology

The research methodology that served as inspiration for the work carried out for this dissertation is Design Science Research. Applied mainly in Engineering and Computer Science, the main goal of design science research is to promote and blossom knowledge in the area of investigation that can help in the design of solutions for problems in the field under analysis.

Hevner *et al.* (2004) developed a Design Science research Framework, that is characterized by seven guidelines, helping the development and built of the research methodology. These guidelines have their foundation in the knowledge and understanding of a design problem and its solutions, that are acquired in the building and application of an artifact, since Design Science is inherently a problem-solving process. The guidelines are the following: Design as an Artifact, Problem Relevance, Design Evaluation, Research Contributions, Research Rigor, Design as a Search Process and Communication of Research. A further analysis of each guideline is made in section 4, Methodology.

## 1.6 Outline

The initial section, **1. Introduction**, has a brief explanation of the background of the study, the description of the problem, and a small introduction to the company where the study took place, along with the research questions, the research methodology and the report outline of this document.

Next, section **2. Literature Review** presents a research and analysis of the existing literature on the three main topics present in this work: Services/Service Design, Data Lineage, and Implementation.

In **3. Problem Characterization,** a deeper analysis of the goals and objectives of the organization and the research was prepared, a small explanation on how Data Lineage inside Euronext is built and an analysis on the different approaches to Service Design.

Following section 3., section **4. Methodology** is divided in a sub-section with a research and comparison of the existing methodologies that can be used for this work to be possible, and a sub-section explaining the methodologies that were chosen.

Section **5. Results** includes an explanation of how the goals of the organization and research were met, and the results of the practical work, and is composed by two sub-sections: Service Design, with the design of Data Lineage as a service, using Multilevel Service Design; and Service Implementation, firstly with general guidelines on Implementation and later with the usage of the guidelines with the actual implementation plan designed with the Data Governance Team.

In the end, section **6. Future Work** presents an analysis on where the whole study can be improved and how that improvement can be done.

# 2 Literature Review

Building on the brief explanation of the main problem presented in the Introduction, a Literature Review of the main subjects was made, in order to guide and support the search for the best solution for the problem described. Accordingly, a literature review of the following themes was made, to develop a deeper knowledge and to support finding the best possible solution for the problem described:

- **Data Lineage**: to understand better the theoretical approach behind the technical part and several approaches to Data Lineage.
- **Service Design**: an explanation on what are Services and the main themes related to them as well as an investigation on different methodologies of Service Design.
- **Implementation**: to understand what Implementation is, and the different areas on which it can be applied and perform an examination of guidelines and frameworks.

## 2.1 Data Lineage

### 2.1.1 Definition

Data Lineage, that can also be known as Data Provenance and Data Pedigree (Bose, 2002), among the several literatures available, has different definitions. However, all the definitions lead to the same core:

*"Data Lineage can be defined as the sources and derivatives of a data set of products or as all the processes and transformations of data from original measurements to current form. It can be described by encompassing data acquisition and compilation methods, conversions, transformations and analysis, along with the assumptions and criteria applied at any stage of the data set life cycle."* (Bose, 2002).

### 2.1.2 Benefits

Understanding the benefits of Data Lineage is a major requirement in order to fully understand what Data Lineage can do to improve the data quality and its use.

Data Lineage can bring several benefits and have a different number of uses when talking about a Lineage or a Provenance system. Simmham (2015) describes different uses for e-commerce, based on a previously performed literature review:

- **Data Quality**: Lineage can be used to estimate data quality and data reliability based on the source data and transformations. It can also provide proof statements on data derivation.
- **Audit Trail**: Provenance can be used to trace the audit trail of data, determine resource usage, and detect errors in data generation.
- **Replication Recipes**: Detailed provenance information can allow repetition of data derivation; help maintain its currency and be a recipe for replication.
- **Attribution**: Pedigree can establish the copyright and ownership of data, enable their citation, and determine liability in case of erroneous data.

Bose and Frew (2005) completed a study on the different benefits of Data Lineage in two components: Data Quality and Scientific Processing. Since one of the main focus, already

mentioned, is how Data Quality benefits from the usage of Data Lineage, the following table (Table 1) presents the benefits that Bose and Frew (2005) identified in their study:

**Table 1 Data Quality Benefits of Data Lineage** (Bose and Frew, 2005)

| **Data Quality Benefits of Data Lineage** | **References** |
|---|---|
| Communicates data quality: suitability, reliability, accuracy, currency, redundancy. | (Lanter, 1991), (Eagan and Ventura, 1993), (Clarke and Clark, 1995), (Buneman, Maier and Widom, 2000) |
| Enhances interpretation, prevents misinterpretation, misuse of environmental data. | (Eagan and Ventura, 1993) |
| Enhances a user's justification for using data. | (Eagan and Ventura, 1993) |
| Reduces possible false sense of data precision. | (Eagan and Ventura, 1993) |
| Facilitates integration of data for regional analysis. | (Eagan and Ventura, 1993) |
| Allows nonexpert data user to understand processing steps. | (Woodruff and Stonebraker, 1997) |
| Communicates processing steps leading to creation of scientific data product. | (Brown and Stonebraker, 1995), (Buneman, Maier and Widom, 2000), (Frew and Bose, 2001) |
| Allows access to sources of materialized relational views; "drill down". | (Cui, Widom and Wiener, 2000), (Buneman, Khanna and Tan, 2001) |
| Allows updates to sources from materialized relational views. | (Cui, Widom and Wiener, 2000) |
| Allows modification of materialized relational view schema. | (Cui, Widom and Wiener, 2000) |
| Enables future generations to use historical data resources. | (Clarke and Clark, 1995) |
| Documents geographical changes from successive updates to a reference cadastral DBMS. | (Spery, Claramunt and Libourel, 2008) |

It is also possible to find definitions and examples of benefits on organizations that are specialized on the study of data. DATAVERSITY, which focuses on producing "educational resources for business and Information Technology professionals on the uses and management of data" (DATAVERSITY, 2020), features an article written by Michelle Knight (DATAVERSITY, 2017) where Data Lineage is analyzed, explaining "What is Data Lineage", "Why keep track of Data Lineage" and "How to create and use Data Lineage in your Business". It is highlighted the "Why" and "How", since the "What" references a very similar previously mentioned definition.

On the "Why" section, the benefits of Data Lineage are explained, mentioning several important concepts. Table 2 mentions the activities mentioned in DATAVERSITY article, a definition of each activity, and the relationship with Data Lineage, also from the DATAVERSITY article.

**Table 2 Activities related to Data Lineage** (DATAVERSITY, 2017)

| Activity | Definition | Relationship with Data Lineage |
|---|---|---|
| **Data Governance** | "Data Governance refers to who holds the decision rights and is held accountable for an organization's decision-making about its data assets" (Khatri and Brown, 2010). | • Data Governance requires Metadata Management;<br>• Needed to ensure Big Data meets business standards;<br>• A Data Lineage solution stitches Metadata together providing "understanding and validation" of data usage and risks that need to be mitigated. |
| **Compliance** | "Certification of confirmation that the doer of an action (such as the writer of an audit report), or the manufacturer or supplier of a product, meets the requirements of accepted practices, legislation, prescribed rules and regulations, specified standards, or the terms of a contact" (Business Dictionary, 2020). | • Different stakeholders (customers, staff members and auditors) need to trust reported data, while quickly responding to "business opportunities and regulatory challenges". |
| **Data Quality** | "Quality of data refers to its ability to satisfy its usage requirements." (Khatri and Brown, 2010).<br>"While data quality has multiple dimensions such as accuracy, timeliness, completeness and credibility, these dimensions are relative and need to be defined in the context of the end use of data" (Khatri and Brown, 2010). | • Data movement, transformation, interpretation and selection through people and processes;<br>• A Data Lineage solution provides the ability to know when "at the end-to-end flow", encompassing when data has been transformed, what it means and how the Data Quality moves from one place to another. |
| **Business Impact Analysis** | "A Business Impact Analysis (BIA) predicts the consequences of disruption of a business function and process and gathers information needed to develop recovery strategies" (Ready.gov, 2015). | • Businesses need to understand how internal departments and users, as well as external customers, share Big Data, especially Master Data and how these data change;<br>• Businesses may wish to upgrade the Data Warehouse and need to know what systems and processes could break doing this. |

On the "How" section, the article mentions some strategies that can be followed to create an effective Data Lineage:

- **Document the Where and How of your Data**: understanding the data that exists, including key business processes and flow between these processes and track where data has moved and how it has changes, in a repeatable, defensible, and speedy manner;

- **Investigate the 5 W's**: finding out "who" is using the data, "what" does it mean, "when" was it captured, "when" is it being used and "why" is it stored and/or used;
- **Understanding Relationships**: relationships between data need to be well understood, including how data originates and moves between people, processes, services, and products;
- **Automation**: technologies build out a reverse tracing methodology and baseline to get comprehensive and end-to-end lineage, identify critical or master data, and use an automated Metadata application to scan and gather Metadata about Data Lineage.

Finally, one of the main benefits of Data Lineage is the possibility to perform a Change Analysis Impact. These two activities relate in the sense that the Change Analysis Impact can "make use of the traceability of data, i.e. where it originates, what its various fields mean and what transformations are needed to perform required analysis" (Hoang, 2011).

In a United States Patent, Subramanian *et al* (2019), presents a "flow diagram of a computerized method of data lineage identification and change impact prediction in a distributed computing environment' that explains and shows how the activity of Impact Analysis can relate with the use of Data Lineage:

1. The metadata associated with the data sources is captured, followed by the determination and identification of the direct and indirect relationships between the data objects (from the data sources) and the metadata captured;
2. The data lineage is generated across the data sources, using the direct and indirect relationships previously identified;
3. Unstructured text is extracted from stored incident tickets;
4. It is generated a multidimensional vector, from the data objects;
5. A change classification model is trained, using the multidimensional vector of the data objects;
6. After the change classification model is created, once a request to change a data object in received, it is determined the change impact score for the change requested and the request is executed once the change impact score is above a predetermined threshold.

### 2.1.3 Systems

It is possible to find in some works a comparison of several systems where Data Lineage can be performed, for example, in the works of Simmham (2015), Glavic and Dittrich (2007) and Bose and Frew (2005).

On the work of Simmham *et al*. (2015) it is possible to find a survey of different works that are related to the different techniques of Data Lineage, that can provide a comprehensive overview of research in this field. In this work, (1) Chimera, (2) myGRID, (3) CMCS, (4) ESSW and (5) Trio are analyzed and compared in different factors: Applied Domain, Workflow Type, Use of Provenance, Subject, Granularity, Representation Scheme, Semantic Info., Storage Repository/Backend, User Overhead, Scalability Addressed and Dissemination.

Is it possible to also encounter in the work of Glavic and Dittrich (2007), a study of different approaches to the provenance of data, from a conceptual point of view. Like Simmham *et al.*, Glavic and Dittrich (2007) gather existing models, categorizing them: (1) Provenance of views in a data warehouse, (2) Trio, (3) Chimera, (4) PRoServ, (5) Copy-paste model, (6) DB-Notes.

They characterize them based on the following factors: World model, Identification, Representation of Transformations, Supported Transformations, Representation of source data, Source definition, Provenance based recreation, Transformation provenance manipulation, Source provenance manipulation, Data model manipulation, Storage strategy manipulation, Query, Storage strategy, Storage model, Propagation and Recording Strategy.

Finally, in the work of Bose and Frew (2005), the authors also mention several systems where Data Lineage can be retrieved, some of them, also mentioned in the two previous works: (1) Chimera, (2) ESSW, (3) GOOSE, (4) Geo-Opera, (5) Tioga, (6) CMCS, (7) MyGrid, (8) S audit facility and (9) TREC.

Bose (2002) created a framework, represented in Figure 1, that explains how Data Lineage can be used:

(1) For scientific workflow,
(2) in conjunction with an appropriate **data model** incorporating lineage,
(3) $t_0$ being the beginning of the process and $t_n$ the end of the process, the system automatically **records** lineage metadata
    a. through a process definition provided at $t_0$,
    b. by receiving metadata as tasks occur from $t_0$ to $t_n$ or
    c. through a log of completed tasks at $t_n$ and
(4) **stores** the lineage metadata in a stand-alone or integrated meta-database (file system of DBMS).
(5) The stores lineage metadata can then be accessed to
    a. answer ad hoc **queries**,
    b. browse or **navigate** (backwards or forwards) through the lineage,
    c. **manipulate** lineage information (for example, query results) to
        i. modify future workflow,
        ii. re-execute past workflow, or
    d. provide lineage reports to communicate data quality, possibly with flexible levels of detail, to other groups of domains.

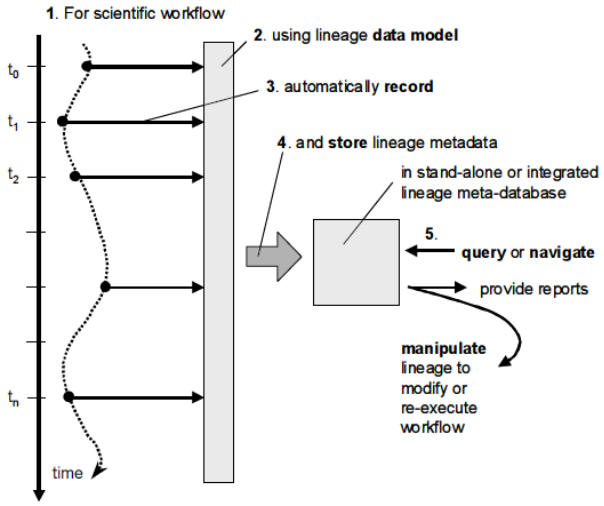


**Figure 1 Framework from Bose for Data Lineage (Bose, 2002)**

On the work of Bose (2002) it is possible to encounter the study of Data Lineage for scientific data, however, Data Lineage can be used within diverse types of data and information. For example, the work of Mwebaze, Mcfarland and Booxhorn (2010) presents an analysis of how Data Lineage can be used in distributed sub-image processing, where the authors "describe a new framework that leverages data lineage and provenance to aid in selective retrieval and processing of data".

As becomes clear from the previous sections, the use of Data Lineage has a more technical and practical approach when it comes to its use in companies, where it is more used as a tool, even with several benefits and characteristics.

## 2.2 Service Design

### 2.2.1 Services

In the world as it is, services are all around us, whether in a restaurant, or in a hospital, everyone has close contact with services in a daily-basis routine. However, the design and presentation of a given service have significant work behind it.

Firstly, there are several definitions of "Service" in the literature, but at its core, services are "economic activities offered by one party to another. Often time-based, performances bring about desired results to recipients, objects or other assets for which purchases have responsibility" (Lovelock and Wirtz, 2011). It is possible to add to this definition that "a service consists primarily of processes and can only be experienced, created or participated in. It is produced and consumed simultaneously, consequently the quality of the service by its nature will therefore be variable" (Alonso-Rasgado, Thompson and Dannemark, 2005).

When it comes to understanding Services and everything that complements them, it is important to realize the change in a paradigm: from a goods-centred view to a service-centred view. In Marketing, before the service-centered focus appeared, the traditional focus was on the production and sale of physical goods. The primary characteristics of the goods-centered view are the following (Vargo and Lusch, 2004):

1. The purpose of economic activity is to make and distribute things that can be sold;
2. To be sold, these things must be embedded with utility and value during the production and distribution processes and must offer to the consumer superior value in relation to competitors' offerings;
3. The firm should set all decision variables at a level that enables it to maximize the profit from the sale of output;
4. For both maximum production control and efficiency, the good should be standardized and produced away from the market;
5. The good can then be inventoried until it is demanded and then delivered to the consumer at a profit.

Prior to 1960, services were only seen in the Marketing literature as an aid to the production and marketing of goods. The service-centered view of marketing "implies that marketing is a continuous series of social and economic processes that is largely focused on operant resources with which the firm is constantly striving to make better value propositions than its competitors" (Vargo and Lusch, 2004).

The service-centred view has the following characteristics (Vargo and Lusch, 2004):

1. Identify or develop core competencies, the fundamental knowledge and skills of an economic entity that represent potential competitive advantage;
2. Identify other entities (potential customers) that could benefit from these competencies;
3. Cultivate relationships that involve the customers in developing customized, competitively compelling value propositions to meet specific needs;
4. Gauge marketplace feedback by analyzing financial performance from exchange to learn how to improve the firm's offering to customers and improve firm performance.

To fully understand the differences from the Traditional Goods-Centred Dominant Logic and the Emerging Service-Centred Dominant Logic, the previously mentioned work by Vargo and Lusch (2004), presents some differences between the two views.

With this, it is crucial to acknowledge the work of Zeithaml, Parasuraman and Berry (1985), when related to the main characteristics of Services. Based on a literature review, the authors identified the main characteristics of services that have been mentioned in different works. The main four characteristics that have been found are:

1. **Intangibility**: intangibility was the main difference cited by the authors, and it means that services cannot be seen, felt, tasted, or touched in the same way goods can be sensed;
2. **Heterogeneity (non-standardization)**: heterogeneity refers to the high variability in the performance of services, this means, that the quality and essence of a service can vary from producer to producer, from customer to customer or even from day to day;
3. **Inseparability of Production and Consumption**: this characteristic refers to the fact that, while goods are first produced, then sold and them consumed, services are first sold and then produced and consumed simultaneously;
4. **Perishability (cannot be inventoried)**: perishability means that services cannot be saved, they are performances that cannot be stored, that can lead to difficulties when synchronizing supply and demand.

Vargo and Lusch (2004) compare the pre 1900 years, where there was a Goods-Centred Model of Exchange, with how the leaders in marketing then continuously moved away from tangible output and toward exchange relationships that involve performing processes and exchanging skills and/or services in which **value is co-created** with the consumer.

This service-centred view of marketing is customer centric, meaning centred on the collaboration with the consumer, rather than embedded in output. The article of Vargo and Lusch (2004), presents eight foundational premises, that help to present the patchwork of the emerging dominant logic:

- $FP_1$ The Application of Specialized Skills and Knowledge Is the Fundamental Unit of Exchange;
- $FP_2$ Indirect Exchange Masks the Fundamental Unit of Exchange;
- $FP_3$ Goods Are Distribution Mechanisms for Service Provision;
- $FP_4$ Knowledge if the Fundamental Source of Competitive Advantage;
- $FP_5$ All Economies are Services Economies;
- $FP_6$ The Customer is Always a Co-producer;
- $FP_7$ The Enterprise Can Only Make Value Propositions;
- $FP_8$ A Service Centered View Is Customer Oriented and Relational.

Focusing on the Foundational Concept number 6, the authors believe that the customer is always the co-producer of value. However, in a more recent paper, Vargo and Lusch (2016), advance that the concept of value co-production has shifted to a concept of value co-creation. This way, they were able to distinguish “co-production”, referring to the creation of the value proposition (design, definition, production, etc), and “value co-creation”, the actions of multiple actors, often unaware of each other, that contribute to each other’s wellbeing, meaning that value is co-created by multiple actors, always including the beneficiary.

With the focus that is given to the customers, Michel, Brown and Gallan (2008) believe that the “focus of innovation should shift towards innovation customer’s value co-creation roles”, having no distinct separation between production and consumption, instead, viewed as

inexorably linked along a continuum of value creation. Based on research, the authors explored three ways to create value with customers:

(1) Creating Smart Offerings,
(2) Changing the Integration of Value and
(3) Reconfiguring Value Constellations.

However, they believe that, to innovate with customers, it is necessary to have two approaches: ”Outside-In” service-logic innovation, which starts with changing customers’ roles, and causes a change in the firm’s value creation, and “Inside-Out” service-logic innovations, which starts by changing the firm’s value creation and then encourages a change in customer’s roles.

### 2.2.2 Service Design

The International Organization for Standardisation [ISO 1991] goes some way to providing a definition of service design. In the Guidelines for Services, Quality Management and Quality System Elements, the process of designing a service is described as one that (Alonso-Rasgado, Thompson and Dannemark, 2005);

- Involves converting the service brief (customer needs) into specification for both the service and its delivery and control, while reflecting the organization’s options (i.e. aims, policies and costs);
- The service specification defines the service to be provided, whereas the service delivery specification defines the means and methods used to deliver the service. The quality control specification defines the procedures for evaluating the controlling the service delivery characteristics;
- Design of the service specification. The service delivery specification and quality control specification are interdependent and interact through the design process. Flow charts are a useful method to depict all activities, relationships, and interdependencies.
- The principle of quality control should be applied to the design process itself.

It is also useful to mention the main service design principles, that provide a deeper insight on the area. Korper *et al.* (2020) wrote a summary of 6 service design principles gathered by Karpen, Gemser and Calabretta (2017):

1. **Human-and meaning-centred**: shifts the focus from technology or products to the experiences and values of the people that will be affected by the service design solution.
2. **Co-creative and inclusive**: provides a space for participation and joint collaboration by multiple actors. It is inclusive in its methods with a goal of identifying the user value of the context of use.
3. **Transformative and betterment-oriented**: considers a multi-part aspect of long-term impact of the service design solution across organizational, societal, and environmental systems.
4. **Emergent and experimental**: problem and solutions aspects of the service design practice are intertwined and emerge together through a service design process that has structure but is often characterized by ambiguity.
5. **Explicative and experientially explicit**: visual representation are essential elements of service design as they bring tangibility to the multiple levels of service consideration. Visualizations are realized through various tools. They are essential for prototyping but are important for other design phases as well.

6. **Holistic and contextual**: service design is systemic and takes a broad perspective when approaching problem-solving. It uses context complexity as a resource in designing meaningful solutions.

With the emergence of Service Design, several methodologies started to rise, with the aim of creating the best way to design a process. When talking about Service Design, it is crucial to understand the type of methodologies that were created, not only using previously existing techniques. Four papers were found in which the authors present a methodology or a set of techniques to Design a Service, and are summarized in Table 3:

**Table 3 Methodologies/Techniques of Service Design**

| **Authors** | **Methodologies and/or Techniques** |
|---|---|
| Chuang (2007) | "Combining the uses of service blueprint and failure analysis in a service company":<br>• **Service Blueprint** ("is a map or flowchart (…) of all transactions constituting the service delivery process");<br>**FMEA** (Failure Modes and Effects Analysis) ("is a reliability analysis tool widely used in the manufacturing sectors (…) to identify, prioritise and eliminate known potential failures, problems and errors"). |
| Holmlid and Evenson (2008) | "Combination of the two high-level categories of methods that can contribute to the advancement of SSME (Service Science, Management and Engineering)":<br>**Human-Centred methods**:<br>• Customer Journeys (the experience of the customer during the service);<br>• Genres ("implicit contract between producer and consumer, directing both the production process and the expectations of the consumer").<br>**Modelling, prototyping and enacting methods:**<br>• Modelling Stakeholders (mapping all the stakeholders and their influence and creating personas);<br>• Modelling activities (what happens, how people act, in what order things happen and coordination of backstage and frontstage activities);<br>• Prototyping ("encompasses the experience as well as the touchpoints");<br>• Enacting ("working with dramaturgic methods allows designers and users to enact or perform service experiences). |
| Morelli (2009) | Service Design divided in three phases:<br>**Analysis and interpretation of the context**:<br>• Mapping the context ("shaped by the socio-cultural frameworks of the actors directly or indirectly involved in the development process")<br>• Profiling the actors ("the list of criteria provides an accurate analytical framework to define actors'").<br>**Ethnographic studies:** |

| | |
|---|---|
| | • Use and interpretation of videos ("ambiguity and open-endedness of interpretation and a high dependency on the participation of actors, recorders, editors and viewers");<br>• Collecting data from users ("cultural probes as a way of encouraging users to record relevant information in photographs, personal diary and through the use of different forms of inspirations");<br>**Design and development tools:**<br>• Service blueprints ("allows for a quantitative description of critical service elements, such as time, logical sequences of actions and processes");<br>• Architecture of the service ("use cases represent a detailed view of a system of interactions at the local level"). |
| Fisk, Patrício and Constantine (2011) | **Studying the Customer Experience** (data collection techniques, such as observation, in-depth interviews, focus groups, usability testing or walkthroughs);<br>**Designing the Service Concept**<br>• Understanding the Value Constellation Experience or VCE ("interactions between the customer and all service organizations");<br>• Customer Value Constellation or CVC ("service offerings and respective interrelationships that enable customers to cocreate their VCE").<br>**Designing the Firm's Service System**<br>• Understanding the Service Experience ("all interactions between a customer a firm's service system");<br>• Service System Architecture (SSA) and Service System Navigation (SSN) ("multiple patterns of navigation across service interfaces").<br>**Designing the Service Encounter**<br>• Understanding the Service Encounter Experience (SEE) ("interactions at a given service interface for a service task");<br>• Service Experience Blueprint (SEB) ("maps the actions of the different participants in the service encounter, both frontstage and backstage"). |

### 2.2.3 Service Blueprints

Inside the area of Service Design, Service Blueprints are one of the sub-themes or techniques that were created, and have an important influence in Service Design (three out of the four methodologies presented in the previous sub-section use Service Blueprint in the process of Service Design).

A Service Blueprint "allows a company to explore all the issues inherent in creating or managing a service" (Shostack, 1984). Shostack (1984) created the technique of Service Blueprint and defined four steps to the process of designing a blueprint:

- **Identifying processes:** mapping the processes that constitute the service;

- **Isolation fail points:** identification of fail points and design of fail-safe processes;
- **Establishing time frame:** establish a standard execution time;
- **Analysing profitability:** establish a time-of-service-execution standard that precludes unprofitable business and maintains productivity.

## 2.3 Implementation

The term Implementation can be described as a "recursive process with steps that are focused on achieving benefits for children, families, provider organizations, human service systems and communities" (Fixsen *et al.*, 2009) and it can be used in several situations, but first it is crucial to understand some different approaches on service implementation.

### 2.3.1 Service Implementation

An approach to Service Implementation is offered by Aarons, Hurlburt and Horwitz, (2011), who designed a Conceptual Model for Service Implementation in Public Service Sectors, with the following steps, each of one them being divided in "Outer Context" and "Inner Context":

- **Exploration:** "The Exploration Phase involves awareness of either an issue that needs attention or of an improved approach to an organizational challenge".
- **Adoption Decision/Preparation:** "The adoption decision is often conceptualized as a one-time event while, in practice, organizations may experiment with an innovation, sometimes intermittently, prior to broader implementation".
- **Active Implementation:** "Although general principles apply to large and small implementation efforts, the scale of the implementation has implications for specific issues in both outer and inner contexts).
- **Sustainment:** "We use the term sustainment to denote the continued use of an innovation in practice".

In another approach, Moeller (2008) focused on designing a framework, called FTU (Facilities, Transformation, Usage), and extending it to Service Dominant Logic (Vargo and Lusch, 2004), seeing it from an implementation perspective. This means that FTU is an implementation framework, "a way to support service provision as a higher-order category of market offering" (Moeller, 2008).

It is also worthwhile to point out that when mentioning Implementation, researchers may be doing it in a more general sense, without having to follow a series of guidelines or a framework. For example, Crews *et al.* (2011) work with the Development and Implementation of a pharmacist-managed clinical pharmacogenetics service and the steps of implementation that are used are based on that context's specific needs and not based on existing guidelines or a framework.

On a different approach, there has been some research associated to how Innovation can relate with other subjects. For example, in the work of Cadwallader *et al.* (2010), the authors focus on understanding and developing a "theoretical model to investigate the complex role of motivation in engaging employee participation in service innovation implementation".

### 2.3.2 Other approaches

In the work of Fixsen *et al.* (2009), two types of guidelines related to Implementation are mentioned: the Stages of Implementation and the Core Implementation Components, the main subject of the work. The Stages of Implementation "can be thought of as components of a tight circle with two-headed arrow from each to every other components" and although they are not the core of the work, consequently, not having a full explanation of each phase, they can serve as guidelines: (1) Exploration, (2) Installation, (3) Initial Implementation, (4) Full Implementation, (5) Innovation and (6) Sustainability.

The Core Implementation Components, also mentioned in the work of Fixsen *et al.* (2009), are related to implementing and sustaining the "Effective Use of Human Service Innovations Such as Evidence-Based Programs", and are the following: (1) Staff Performance Evaluation, (2) Decision Support Data Systems, (3) Facilitate Administrative Supports, (4) Systems Interventions, (5) Recruitment and Selection, (6) Preservice Training and (7) Consultation & Coaching.

Chamberlain, Brown and Saldana (2011) developed the Stages of Implementation Completion, or "SIC", based on their experience of working Multidimensional Treatment Foster Care (MTFC), an evidence-based program that is an alternative to residential care for children and adolescents. In this work, SIC is used to examine the implementation of MTFC. They developed SIC with the following Phases and Stages:

1. **Pre-Implementation:**
    i. Engagement;
    ii. Consideration of Feasibility;
    iii. Readiness Planning;
2. **Implementation:**
    i. Staff Hired and Trained;
    ii. Adherence Monitoring Processes in place;
    iii. Services and Consultation Begin;
    iv. Ongoing Services, Consultation, Fidelity Monitoring and Feedback;
3. **Sustainability**:
    i. Competency.

In a different approach, Spillane, Reiser and Reimer (2002), developed a "cognitive framework to characterize sense-making in the implementation process", that can be mainly used in Policy Implementation. This way, the authors divided the framework in three stages: (1) Individual cognition, (2) Situated Cognition and (3) Roles of representations. The development of this framework comes from the interaction between "their [implementing agents] existing cognitive structures (including knowledge, beliefs and attitudes), their situation and the policy signals".

On a more general vision, Nilsen (2015) performed an analysis of the different theories, models and frameworks used in implementation science, and divided them into five categories, and providing for each one several references (it is advised to see the original work more in depth):

- **Process Models**: "Specify steps (stages, phases) in the process of translating research into practice, including the implementation and use of research. The aim of process models is to describe and/or guide the process of translating research into practice. (…)";

- **Determinant Frameworks**: “Specify types (also known as classes or domains) of determinants and individual determinants, which act as barriers and enablers (independent variables) that influence implementation outcomes (dependent variables). (…)”;
- **Classic Theories**: “Theories that originate from fields external to implementation science, e.g. psychology, sociology and organizational theory, which can be applied to provide understanding and/or explanation of aspects of implementation”.
- **Implementation Theories**: “Theories that have been developed by implementation researchers (from scratch or by adapting existing theories and concepts) to provide understanding and/or explanation of aspects of implementation.”;
- **Evaluation Frameworks**: “Specify aspects of implementation that could be evaluated to determine implementation success”.

# 3 Problem Characterization

This section aims at better understanding and analysing the problem addressed in this study: first, it provides more context on the problem outlined in the Introduction; then it describes Financial Services broadly and provides more information about Euronext Technologies, the context for this study; after that, the objectives for this research are defined; then, the work done before the implementation of Data Lineage, that is, the mapping of the Data Lineage, with the visual representation, is presented; and finally, different approaches to Service Design are introduced, to understand how Data Lineage can be designed with it.

## 3.1 Problem Context

The activities related to Data Governance that are integrated in a company can depend on multiple factors, but there are some main activities that best represent the contributions that Data Governance can bring to a company (Search Data Management, 2020), represented also on Figure 2:

> *"Data Governance relies on **data stewards** to implement policies in organizations, a central goal being better **data quality**. Often, a key element in data governance implementations is **master data management.** Like data governance itself, MDM continues to encounter new **use cases** as data use widens."*

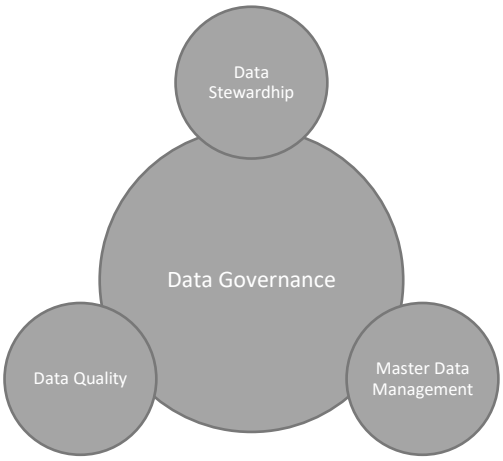


**Figure 2 Data Governance activities (Search Data Management, 2020)**

According to Khatri and Brown (2010), the main domains of Data Governance are:

- **Data Principles**: clarifying the role of data as an asset;
- **Data Quality**: establishing the requirements of intended use of data;
- **Metadata**: establishing the semantics or "content" of data so that it is interpretable by the users;
- **Data Access**: specifying access requirements of data;
- **Data Lifecycle**: determining the definition, production, retentions, and retirement of data.

Data Lineage is one of the main activities that can be performed in Data Governance. This way, it is important when a Data Lineage activity is implemented in a company, that it is understood and known by every department, and all the benefits that Data Lineage can bring to the company are perceived.

As in the case of Euronext, a problem arises when Data Lineage is already being performed but does not have the visibility that is necessary in order for the all departments and for all company to benefit, despite all the efforts that weredone by Data Governance, such as, demos and presentations about Data Lineage.

Unfortunately, this becomes a problem with two sides: one for the department that is currently mapping the Data Lineage, because its work is not being seen by the other departments; the other for the other departments, since Data Lineage could already be helping and solving some problems, but due to the lack of visibility, that is not possible. This required creating a plan of implementation for Data Lineage.

## 3.2 Financial Services and Euronext Technologies

According to the IMF (International Monetary Fund), (Asmundson, 2012), to fully understand what is a financial service, it is necessary first to analyse the difference between: a good, "something tangible that lasts, whether for a long or short time", and a service, "a task that someone performs for you". The financial service is not the financial good, but the "process of acquiring the financial good, (…) it involves the transaction required to obtain the financial good".

As such, the financial sector "covers many different types of transactions in such areas such as real estate, consumer finance, banking, and insurance (…), a broad spectrum of investment funding, including securities.". Issue securities are an activity where "providers help borrowers raise funds by selling shares in businesses or issuing bonds" (Asmundson, 2012).

Based on the information of Euronext's website (Euronext, 2020) and LinkedIn profile, Euronext is the leading pan-European exchange, with a global reach and local presence, covering Belgium, France, Ireland, The Netherlands, Norway, Portugal and the UK. Euronext operates a regulated and transparent equity and derivatives market and it is the largest centre for debt and funds. The total product offering includes Equities, Foreign Exchange (FX) expansion platform, Exchange Traded Funds, Warrants & Certificates, Bonds, Derivatives, Commodities, and Indices.

Euronext is divided in three components: **sales offices** on three continents, from which the international investors and trading members can take benefit, the **regulated markets** in Amsterdam, Brussels, Dublin, Lisbon and Paris and their **technology centres**, where Euronext develop market leading solutions to enable trading in an evolving, highly demanding landscape.

In Portugal, Euronext is present in two cities: in Lisbon, with the Stock Exchange (that enables the companies to "raise capital to finance operations, enhance its visibility and credibility and provide liquidity to its shareholders while retaining control" (Euronext, 2020)) (listing and trading), and in Porto, with Interbolsa (post-trading), integrated in 2000 and Euronext Technologies (Tech Centre), integrated in 2017. Since 2002 the Lisbon exchange has been part of Euronext, which is the preferred listing venue for Portuguese issuers.

The work that it is being done in this dissertation has its origin in Porto, in the Tech Centre, more specifically inside Data Services – Data Governance.

It is important to understand what are the main activities performed by the Data Governance and their benefits. The Data Governance department has the capabilities of handling and managing metadata, data sources and master data management, compliance checking and matters of both regulatory and legislative impact including matters relating, but not limited to, data privacy (more specially, data masking), data governance and data management best practices. Data Governance plays a very important role inside Euronext, since it works with very useful tools that can improve and ease the use of data and assures compliance of the organization.

### 3.3 Organization Objectives

The Data Lineage initiative inside Data Governance started as part of the compliance, regulatory and GDPR program in 2018. However, since then, the main difficulty has been to raise awareness of it to other data stakeholders, of how this framework could be used and leveraged for their needs (Software Development Life Cycle, etc.). So, it became crucial as a Data Governance strategy, to establish a Data Lineage service to involve all relevant stakeholders to participate with inputs to the roadmap and as service clients.

The department of Data Governance has a very important role inside Euronext Technologies, but it was difficult to show that to the rest of the company. As such, the main goal of this work was to have Data Lineage implemented as a service in Euronext, with its creation and implementation building on three general steps:

- Creation of the Data Lineage mappings;
- Design of the service;
- Implementation of the service.

For the first step, the creation of the Data Lineage mappings, the organization already had full guidelines for the whole process, using a Metadata Management platform, Talend Data Catalog (Talend, 2020). This process is fully explained later, since it is one of the tasks that was performed. However, it can be advanced that the process uses two components of Talend:

- Talend Big Data Integration, for performing an analysis of the ETL (Extract, Transform, Load) jobs;
- Talend Data Catalog for the mapping of the Data Lineage.

For the second step, the design of the service, several methodologies are analysed and compared in the section Methodology, to support selecting the appropriate solution for the problem of the organization.

For the final step, the implementation of the service, along with the team of Data Governance, several steps and guidelines are proposed to implement the service according with the rules of the organization and the guidelines found in the literature.

Based on the Literature Review that was done, it is important for the company and for the work itself to understand how a Service Design methodology can be used with an activity like Data Lineage, due to the differences that exists between Data Lineage and other regular services, for example, a new service in a retail company or in a pharmaceutical company. This requires a different approach on using Service Design techniques with a different type of service.

## 3.4 Data Lineage

Before starting the design of Data Lineage as a process, the work of mapping Data Lineage was also done. The next sub-sections explain how this work was done, taking in consideration the steps in Euronext.

### 3.4.1 Data Lineage Mapping

Euronext deals with a large number and diversity of data sources and structures, for example, databases and applications of the organization. Consequently, the data architecture of Euronext is very complex, and the Data Office department, composed by Data Governance and Analysis, Quality and Reporting, has the main goal of managing that data, in order to be easier and simpler to explore and analyse all the data and to help the different departments (from Business to IT) to understand all the data and to have a central repository of data.

For that, Data Governance uses a Metadata Management application, Talend, that offers a large number of solutions related to Data Management. For this work, the solutions involved are Talend Data Catalog and Talend Big Data. The main goal of Talend Data Catalog in Euronext is to be a central repository of all the data and metadata, with the following general features (Talend, 2020):

- Faceted search, data sampling, semantic discovery, categorizing and auto-profiling;
- Social curation with data tagging, comments, review, promotion, certification;
- Data relationship discovery and certification;
- Automatic discovery of the data lake and other data stores.

Another set of important features includes the Management and Monitoring of Talend Data Catalog (Talend, 2020):

- Metadata documentation and end-to-end data lineage;
- Active/passive failover switching;
- Impact analysis and change alerts;
- Version control system;
- Approval workflows for business glossary authoring;
- Customizable UI and REST API.

Understanding that it is possible to perform the Data Lineage mappings using the current tool, the next step is to understand the modifications that the data goes through. For that, Talend Big Data, where the ETL Jobs (that stands for Extract, Transform, Load) are done by the Data Developers, is used. These ETL Jobs are defined by three database functions (Extract, Transform and Load), usually in SQL (Structure Query Language) language, that enable passing data from one database to another, or from one file to a database and vice-versa. This is where the source of the data, the transformations and the final destination of the data are discovered.

Data Lineage is thus performed in three general steps:

1. Discovering and understanding the ETL Jobs (Talend Big Data) or analysing documentation that contains the flow of data;
2. Mapping the transformations, creating the Data Lineage, manually (Talend Data Catalog);
3. Analyse the graphical representation of one ETL Job (Talend Data Catalog).

These three steps are detailed next and illustrated in the following images, which for confidentiality motives, are a simple representation of the type of ETL jobs that exists currently at Euronext.

```
1   insert into file1 (id, full_name, code, begin_date, end_date)
2   select
3   db1.id,
4   (db1.first_name + db1.last_name) as full_name,
5   db1.code,
6   to_char(db1.begindate, 'yyyy-MM-dd') as begin_date
7   to_char(db1.enddate, 'yyyy-MM-dd') as end_date
8   from db1
```

**Figure 3 Example of SQL query for Data Lineage**

Step 1 is illustrated in figure 3, where it is possible to see a new file that is created (file1) and a database, already existing (db1). This SQL job represents a file that was created and needs to be sent to the clients of Euronext with the information that is already on the db1. It is known that not all people usually know how to read SQL code, so, in order to better understand the operations that are being done, it is possible to use also figure 4, the manual mapping for Data Lineage, to better understand what operations are happening.

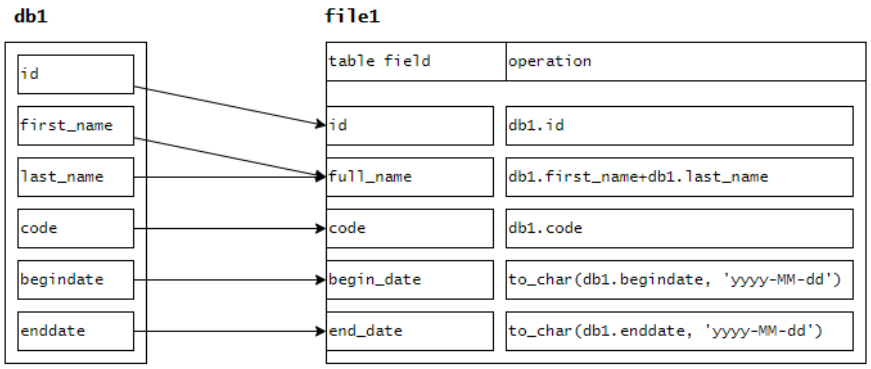


**Figure 4 Example of Mapping of Data Lineage**

From the SQL query, in step 2 (illustrated on figure 4), the responsible (from Data Governance) develops manually the links and connections, from the database to the file. In this step, the information that is described, is the same as in step 1, the difference being that, no operation is done, it consists of simple description of the SQL query. In this case, it is easier to understand the operations of the SQL:

- the field **id** is being mapped directly from db1 to file1;
- in db1, the **first_name** and **last_name** are being put together in a single field **full_name**, so, instead of having two fields for the full name, only one exists;
- the field **code** is being mapped directly from db1 to file1;
- the fields **begindate** and **endate** have a "to_char" operation, meaning, that, based on the format that exists in db1, to file1 the dates are going to be in the format "yyyy-MM-dd".

In step 3, a graphical representation is automatically generated by Talend Data Catalog, as displayed in figure 5. In this case, it is possible to understand the links between db1 and file1, however, it is also possible to understand that the data that was on db1, comes from a source_db, the main database where the data in loaded from. However, in the connection between source_db and db1, not every field is needed, so, there is already a filter. To see the full operation, it is necessary to select each field individually. Here, it is presented the source, the transformations and destination and the whole flow of the data. Each of these operations is performed manually, so, depending on the size of the database or the job, it may be required a large amount of time.

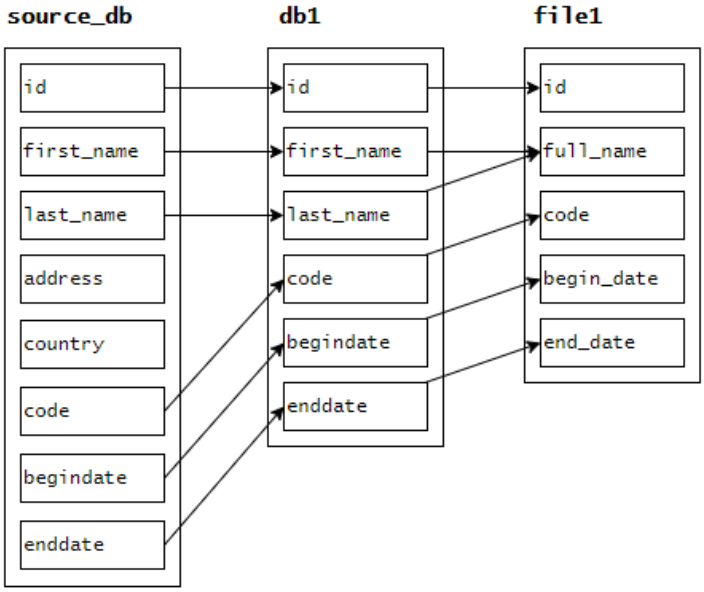


**Figure 5 Example of full diagram of Data Lineage**

### 3.5 Different approaches to Service Design

In order to understand how Service Design can be applied in a case such as Data Lineage, it is relevant to examine different approaches to its application:

- In the work of Karwan and Markland (2006), Service Design and Information Technology techniques are used in order to improve the delivery and productivity in Public Sector Operations, using the case of South Carolina DMV (Department of Motor Vehicles). The authors adapted a framework of Service Design (previously more focused on the private sector), to ensure that it would make sense to the public sector;
- Thomas *et al.* (2009) use Car Parking Services, to explore Service Blueprints (one of the techniques used in Service Design, previously described in the Literature Review) and to understand how this technique "can be applied in a situation with three key actors, all with different motives and wishes";
- With a similar approach to the Car Parking Services case study, the case study by Wang, Lee and Trappey (2017) incorporates TRIZ (Theory of Inventive Problem Solving) and QFD (Quality Function Deployment) with Service Design Blueprints, for a meal

ordering system. The combination of the three techniques "can improve performance and competitiveness of service innovation";

- With a different approach, Kim *et al.* (2018) examined the use of customer behaviour data in service design. The case study is on eco-driving service design using the behaviour data of bus drivers. With that, the authors developed six iterative steps taking in consideration the characteristics of customer behaviour data;
- The next case study uses a three step Service Design methodology (Exploration, Ideation and Reflection and Implementation and Evaluation) in order to develop a national Health Information Systems (Grenha Teixeira, Pinho and Patrício, 2019);
- In another approach, Rossmann and Young (2020) worked on Virtual Reality Spaces and Services (which the authors called "Cyberdiscovery Space"), focusing on a technique from Service Design, Participatory Workshops, that "build empathy for the various people involved in delivering and using the service, as well to generate operational evidence that supports decision-making to improve the service". This technique was used to "generate a shared mission and vision about the purpose and use of the space".

It is possible to understand by the six previous examples that there can be several approaches to the application of Service Design techniques to different types of services. Using a methodology of Service Design to design Data Lineage as a service inside a Financial Services company also requires its specific approach to apply Service Design. This difference on the type of approach comes from the fact that Data Lineage is designed as an Internal Service, so, the "service provider" and the "customers" are included in the same company: Data Governance being the "service provider" and the other departments the "customer".

# 4 Methodology

## 4.1 Existing approaches

### 4.1.1 Data Collection Methods

There are two major types of Data Collection methods, based on the type of Research (Warren, 2020):

- **Qualitative Research:** qualitative research helps to understand people's perceptions and experiences, i.e., is used to understand the human perception; it is often used to formulate theories and identify patterns, so, it is useful for exploratory research.
    - **Qualitative Data Collection and Analysis**: the qualitative research approach uses a small sample size, due to the amount of information and data that can be extracted from one person (in interviews, focus groups or observation).
- **Quantitative Research:** quantitative research is often used to test an already theorised hypothesis, where the data is analysed using statistical analysis, for example, averages or relationship between two or more variables.
    - **Quantitative Data Collection and Analysis:** quantitative research gathers data from a large sample due to the limited data and information extracted from one person (in surveys or questionnaires, with close-ended questions).

Based on the objectives for this work, the method that is going to be used is the Qualitative Research method. It is important to understand what the different departments of Euronext think on subjects that are not quantifiable, so, a more personal method should be used, such as the interviews, to lead to conversations to other topics and to discover new ideas.

### 4.1.2 Service Design Methodology

In the Literature Review section, four methodologies and/or techniques of Service Design were presented and explained:

- A three-phase method (Morelli, 2009):
    - **Analysis and interpretation of the context;**
    - **Ethnographic studies;**
    - **Design and development tools;**
- The combination of service blueprint and failure analysis in a service company: **Service Blueprint** and **FMEA** (Chuang, 2007);
- The combination of the two high-level categories of methods that can contribute to the advancement of SSME (Holmlid and Evenson, 2008):
    - **Human-Centred methods**;
    - **Modelling, prototyping and enacting methods**;
- Multilevel Service Design (Patrício *et al.*, 2011):
    - **Studying the Customer Experience**;
    - **Designing the Service Concept;**
    - **Designing the Firm's Service System;**
    - **Designing the Service Encounter;**

Based on the steps of each methodology or technique, the method proposed by Chuang (2007) does not fit the objectives of this work because it does not give a global view of the whole service. It can be interesting to understand how the two methods work together (Service Blueprint and Failure Modes and Effects Analysis) in a non-typical usage of each method, but it does not fully match the objectives of this work.

The other three methods, by Holmlid and Evenson (2008), Morelli (2009) and Fisk, Patrício and Constantine (2011) are very similar in the way that they divide the service design in different steps. Since the methods that have the best division and a "step by step" methodology are by Morelli (2009) and Fisk, Patrício and Constantine (2011), it was decided not to use the methodology by Holmlid and Evenson (2008).

The two remaining methodologies have a similar way of approaching service design:

- **Gathering of data:**
  - Ethnographic studies (Morelli, 2009);
  - Studying the Customer Experience (Patrício *et al.*, 2011).
- **Understanding the context:**
  - Mapping the context and profiling the actors (Morelli, 2009)
  - The Value Customer Experience and Customer Value Constellation (Patrício *et al.*, 2011).
- **Mapping the service:**
  - Service Blueprint and Architecture of the Service (Morelli, 2009);
  - The Service System Navigation, Service System Architecture and Service Experience Blueprint (Patrício *et al.*, 2011).

However, the methodology created by Fisk, Patrício and Constantine (2011) divides it in four steps, that all relate to each other, understanding every level of the design of service, with, for each step, a global understanding of the level (with the design of the customer experience) and then a practical design (with the Customer Value Constellation, Service System Navigation and Architecture and the Service Experience Blueprint), with each tool being useful for the design of the service. So, in order to have the best possibility of designing a service that suits the objectives of this work, without ignoring any important parts, Multilevel Service Design by Fisk, Patrício and Constantine (2011) was the chosen methodology.

## 4.2 Selected Methods

### 4.2.1 Research Methodology

In order to characterize the subject of study of this dissertation, an analysis has been made having in account the seven guidelines (Hevner *et al.*, 2004), shown in Table 4 and previously explained in the Introduction:

**Table 4 Design Research guidelines** (Hevner *et al.*, 2004)

| Guideline | Description |
|---|---|
| *1: Design as an Artifact* | Produce a viable artifact in the form of a construct, a model, a method, or an instantiation. |

| | |
|---|---|
| *2: Problem Relevance* | Develop technology-based solutions to important and relevant business problems. |
| *3: Design Evaluation* | The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods. |
| *4: Research Contributions* | Provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies. |
| *5: Research Rigor* | Application of rigorous methods in both the construction and evaluation of the design artifact. |
| *6: Design as a Search Process* | Utilizing available means to reach desired ends while satisfying laws in the problem environment. |
| *7: Communication of Research* | Presented effectively both to technology-oriented as well as management-oriented audiences. |

*Design as an Artifact:* The artifact that is being developed is the Data Lineage, a graphical representation that includes the origin of the data, when it was collected/integrated, what applications or databases does this data pass through, how did the data change, and what data sources have confidential information or personal data implications. Data Lineage is going to be integrated as a Service, using Multilevel Service Design.

*Problem Relevance:* The number and diversity of data sources and structures, such as databases and application, make the use of Data Lineage a great opportunity to improve Data Quality. Data Lineage is going to be used as a tool for analyzing the impacts of data changes and as a result increase its quality and reliability, making any corruption or consistency failure evident and easily identifiable. On the other hand, it is going to help every department have a global vision of data in the company.

*Design Evaluation:* Interviews and feedback, together with a prototype, are the techniques used to evaluate the design of Data Lineage. Since Data Lineage is going to be a Service, a Multilevel Service Design methodology is going to be used to develop the Data Lineage.

*Research Contributions:* The use of Multilevel Service Design with a different approach, in an internal service like Data Lineage.

*Research Rigor:* Relevant research was carried out in the field of Service Design and Data Lineage, analyzing past research and works from both areas.

*Design as a Search Process:* Taking into account the problem, a further study was made to understand why Data Lineage was the best solution. Along with Data Lineage, it was also necessary to choose a Service Design methodology. Multilevel Service Design was chosen due to the different levels that are integrated, being able to provide a more global vision of the whole service.

*Communication of Research:* Data Lineage has already the main goal to be used in every department, whether they are more technology-oriented or management-oriented departments,

for them to have a global vision of all the data. This way, this work was developed with the goal that any person can understand every chapter and the work that was done.

### 4.2.2 Qualitative Research

Based on the two types of Research previously mentioned, and each Data Collection methods, the Qualitative Research was the chosen one. This method is going to be used in Step 1 of the Multilevel Service Design, described in the following sub-section. For the Data Collection, the main technique used was open-ended interviews.

There are three types of interviews (Berg and Lune, 2017):

- **Standardized Interviews:** most formally structured, no deviations from question order, wording of each question asked exactly as written, no adjusting of level of language, no clarification or answering of questions about the interview, no additional questions may be added, similar in format to a pencil-and-paper survey;
- **Semistandardized Interviews:** more or less structured, questions may be reordered during the interview, wording of questions flexible, level of language may be adjusted, interviewer may answer questions and make clarifications, interviewer may add or delete probes to interview between subsequent subjects;
- **Unstandardized Interviews:** completely unstructured, no set order to any questions, no set wording to any questions, level of language may be adjusted, interviewer may answer questions and make clarifications, interviewer may add or delete questions between interviews.

The type of Interviews that was used was the Unstandardized type, and later it is explained more in depth the choice.

Various interviews were made with the Data Governance team to collect information about needs, main goals, and objectives, and how the use of Data to be used in Euronext is regarded.

The open-ended interviews were made with a focus on understanding how the different departments and main stakeholders of Data Lineage perceived it, and what were its main benefits.

#### 4.2.2.1 Interviews Structure

##### MAIN STAKEHOLDERS

The interviews were focused on the different departments, mainly the stakeholders and users of Data Lineage. A demo presentation was made to each interviewee on what is Data Lineage and how it works, based on a POC (Proof of Concept) and several mappings of Data Lineage that had already been done as a test. From the knowledge that was gained with this presentation, a set of questions were developed in line with the objectives of the work.

Related to the type of interview (explained previously), the type of interview that was used was the Unstandardized Interview. A set of questions were prepared, but they could vary depending on the topic of conversation. One of the concerns in these interviews was to leave the interviewees as comfortable as possible, explaining that the interviews were informal and even encouraging conversation and discussion.

Due to the small number of possible users, eight interviews were made, covering the following departments at Euronext (in each department, one or more employees may have been interviewed): Enterprise Architecture, Business Analysis, Advanced Data Services, ABS (Application and Business Services), Development Factory and Data Office. Another four interviews were made to stakeholders, with a different structure, as they took place later in the work.

The questions are presented in the following table (Table 5):

**Table 5 Interview questions**

| # | Questions | Explanation |
|---|---|---|
| 1 | **Job Description** | This question serves as a context question, and, if further needed, to compare the types of answers based on the type of work done in Euronext. |
| 2 | **Impact on job** | Based on the daily basis job that the interviewee performs, how the implementation of Data Lineage can impact it. |
| 3 | **Benefits** | The main benefits that Data Lineage can bring to Euronext, not only on the interviewee's department, but also, in the company as a whole. |
| 4 | **Problems where it can help** | Problems the interviewee encountered that Data Lineage could have helped to solve or solve faster or easier. |
| 5 | **Constraints/Problems** | The constraints or problems that Data Lineage can face, not only during its implementation, but also during its use. |
| 6 | **Types of projects** | Based on the interviewee knowledge of the types of projects inside Euronext, in what types Data Lineage can be used. |
| 7 | **Examples of projects** | Old or current projects that Data Lineage could bring benefits to or can impact. |

The main goal of these interviews, as previously mentioned, was to understand what the perception of the employees regarding Data Lineage was. In particular, the implementation of the service required gathering information on the benefits of Data Lineage, in order for Data Governance to be able to provide evidence that Data Lineage bring added value to Euronext.

#### DATA GOVERNANCE

At Euronext, there are two main departments that work with Data on a daily basis: Data Services and Data Office. In order to create the service, it was first necessary to also understand the main activities that were already performed.

Accordingly, unstructured interviews were made to the managers in charge of Data Services and Data Office. The results of these interviews were used mainly with Multilevel Service Design, so they will be presented later jointly with the design of the service. These interviews with the managers and different employees of Data Services and Data Office were made over the work time, with different goals, either to understand the next steps of the work, or to validate previous work.

Due to the unstructured nature of these interviews, no script was created, and all interviews focused on a more natural conversation, focusing on the main information needed. The application of Multilevel Service Design involved these interviews and the analysis of documentation gathered in the company.

### 4.2.3 Multilevel Service Design

Multilevel Service Design or MSD is a methodology first introduced in 2011 by Lia Patrício, Raymond P. Fisk, João Falcão e Cunha and Larry Constantine. This methodology is presented as a new interdisciplinary method for designing complex service systems, that is differentiated from other methodologies due to the use of three different areas: Service Development, Interaction Design and Service Design. With the multilevel perspective, MSD offers a holistic view, from the service concept level to the multi-interface service system level and to each service encounter. The different levels provide different views of the service offering that can be used by different members of the design team and different decision makers.

It divides the designing of a new service in three hierarchical levels: (1) Designing the service concept with the customer value constellation of service offerings for the value constellation experience, (2) Designing the service system, comprising its architecture and navigation, for the service experience and (3) Designing each service encounter with the Service Experience Blueprint for the service encounter experience. Each level is divided into two phases: (1) Understanding the Customer Experience and (2) Designing the Service Offering. The methodology is explained in **Figure 6.**

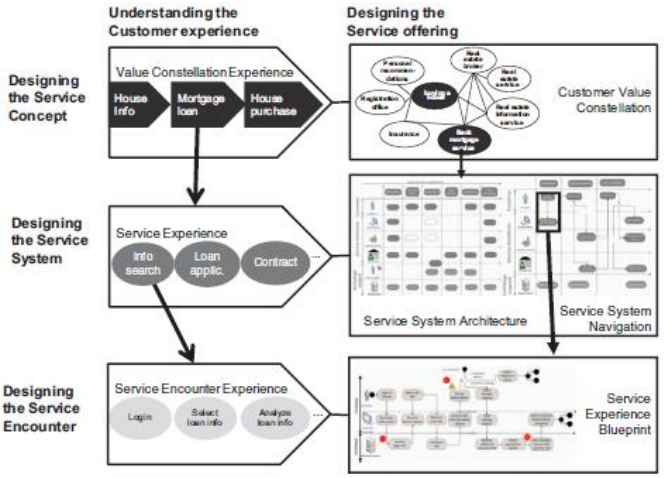


**Figure 6 Components models of Multilevel Service Design (Patrício *et al.*, 2011)**

#### STEP 1: STUDYING THE CUSTOMER EXPERIENCE

The first step of the methodology is **Studying the Customer Experience,** through an in-depth study of the customer experience at its different levels (Service Concept, Service System and Service Encounter). This step involves data collection techniques, focusing on qualitative studies, such as observation, in-depth interviews, focus groups, usability testing or walkthroughs.

#### STEP 2: DESIGNING THE SERVICE CONCEPT

MSD defines the service concept as the firm's positioning in the customer value constellation (CVC), including the services offered and the links and partnerships established with other organizations in the network to enhance the firm's value proposition.

The first part is **Understanding the Value Constellation Experience (VCE)**, that is cocreated through the interactions between the customer and all service organizations that enable a given customer activity. Through the understanding of the VCE, it is possible to **design the Service**

**Concept through the Customer Value Constellation (CVC).** The CVC represents the set of service offerings and respective interrelationships that enable customers to cocreate their value constellation experience for a given customer activity.

#### STEP 3: DESIGNING THE FIRM'S SERVICE SYSTEM

In step 3, the firm designs its service system to enhance the service experience, according to the firm's positioning in the CVC. The MSD method adopts the concept by Maglio et al. (2009) of service system as a configuration of people, technologies, and other resources.

The first part is **Understanding the Service Experience**, that is cocreated through all interactions between a customers and a firm's service system to accomplish a given service activity and includes all the different service encounters with the firm across different service interfaces. Based on understanding the service experience, it is possible to **design the Service System through the Service System Architecture (SSA) and the Service System Navigation (SSN).** This includes three components: (1) for each service task, the service system should offer an interface mix that enables customers to choose their preferred service interface, (2) the service system should enable customers to smoothly navigate across service interfaces through the different tasks of the service activity and (3) instead of replicating every offering in every service interface, the service system should enhance the service experience while contributing to efficient resource allocation among interfaces.

The SSA defines the structure of the service system, providing an integrated view of the multi-interface offering and support processes across the different tasks of the service experience. While the SSA provides a static view, the SSN offers a dynamic view of the service system. Based on the matrix developed in the SSA, the SSN maps the alternative paths customers may take across different service encounters forming the service experience.

#### STEP 4: DESIGNING THE SERVICE ENCOUNTER

The final step is based on defining the moment of interaction between the customer and the firm and may take place in multiple interfaces.

The first part is **Understanding the Service Encounter Experience (SEE)**, that is cocreated through customer interactions at a given service interface for a service task. Through the SEE, it is possible to **design the Service Encounter with the Service Experience Blueprint (SEB) diagram**. This is used to design each concrete service encounter and maps the actions of the different participants in the service encounter, both frontstage and backstage.

# 5 Results

This section presents the practical work that was accomplished, taking in consideration the previously outlined objectives and research background. It is divided in two sub-sections: **Multilevel Service Design**, on how the service (Data Lineage) was researched and mapped inside the company, using Multilevel Service Design; and **Implementation and Validation**, explaining the process of implementation and the steps that were taken, along with a part explaining the process of validation within the company where this study took place.

It is important also to mention that the design of the service required a prior Data Lineage manual mapping with the steps previous explained in the Data Lineage mappings sub-section, inside Problem Characterization section.

## 5.1 Multilevel Service Design

As previously mentioned, the methodology used to design the new service was Multilevel Service Design, which is divided into four phases. The use of Data Lineage is the new service that is designed and implemented, and all the research done during the practical work is based on the information discovered in the Literature Review and the information that was collected from Euronext.

### 5.1.1 Studying the Customer Experience

For the first step of the MSD, Studying the Customer Experience, one main type of Qualitative Research was used, Interviews. These interviews were divided in two groups of interviews:

- The first group of interviews was done to the different stakeholders and potential users of Data Lineage, in order to understand its benefits;
- The second group of interviews was focused on the members of the Data Governance team and current users to understand the whole Service Experience.

#### 5.1.1.1 Interviews

In the first phase, Studying the Customer Experience, there was a need to divide the work in two parts: the first part focused on understanding the perception of the different stakeholders (in different departments) of the benefits and problems of Data Lineage, focusing also on how it could be used; the second part consisted of several interviews with the members of the Data departments, to understand the main activities performed related to Data, and to understand the experience of the other departments when reaching out to the Data departments.

In the following sub-section, **Results**, it is described the results of the first set of interviews (to the different stakeholders of Data Lineage). It is thus more focused on the benefits and different uses of Data Lineage. The second set of interviews is more focused on designing and understanding the Levels of Customer Design.

#### 5.1.1.2 Results

This section is going to be divided into two main parts: (1) the results of the interviews with the main stakeholders, that, although not contributing to the design of the service, are a very effective way to show to the organization what Data Lineage is; (2) the design of the Customer

Experience, that is divided in three main stages, as previously mentioned in the introduction to Multilevel Service Design, under the Methodology section: Value Constellation Experience, Service Experience and Service Encounter Experience.

#### INTERVIEWS

It was necessary to carry out an analysis on the different answers, due to the extensive information that was gathered (Table 6).

**Table 6 Summary of Interview answers**

| Questions | Main topics |
|---|---|
| **Impact on job** | • **Access, prepare and identify a change in the data chain:**<br>With a modification on a file that is not planned (a consequence of a change upstream or downward of the data chain);<br>• **Help the onboarding**<br>Joining the architecture diagrams and the Data Lineage diagrams for a better understanding of Euronext data architecture;<br>• **Better requirements and help in the first use cases**:<br>With a better understanding of the whole systems and their impacts since the beginning of each project, it's easier to not miss any important data. |
| **Benefits** | • **Improve the Quality of the Data**<br>o Consistency of the data;<br>o Identifying and performing Impact Analysis;<br>o Centralized application that uses the right terminology;<br>o Understanding the flow of systems and applications.<br>• **Have a more robust testing strategy;**<br>• **Trigger less expenses for each project;**<br>• **Good addition to the Project Management Policy;**<br>• **Prevent problems and anticipate bugs.** |
| **Problems to solve** | • Troubleshooting;<br>• Incident and Change Management;<br>• Lack and Duplication of Information;<br>• Prevent emergencies at the end of the projects. |
| **Constraints/Problems** | • Understanding all the data;<br>• Complexity and heterogeneity;<br>• Costly (manpower and time);<br>• Big number of changes;<br>• Information can be outdated;<br>• Implement in the correct level; |

| | |
|---|---|
| | • Culture of the company;<br>• Communication must be clear;<br>• Lack of documentation;<br>• Administrator Management. |
| **Types of project** | • **Projects that are relying heavily on data:**<br>The Data Chain needs to be clean and robust.<br>• **IT Projects:**<br>Make the roadmap and strategy consistent and perform the data analysis in the beginning of the data project.<br>• **Projects for regulatory reporting:**<br>Confirm that the data is stable and consistent. |

### 5.1.2 Levels of Customer Experience

The Customer Experience is defined by the experience of the customer through the three levels of service design: Service Concept, with the understanding of the Value Constellation Experience, Service System, with the understanding of the Service Experience and the Service Encounter, with the Service Encounter Experience. Based on the results of the interviews, it was possible to understand the experience of the different departments regarding the work with Data Governance (Figure 7). In the different phases, each level is going to be explained.

#### VALUE CONSTELLATION EXPERIENCE

To understand the Value Constellation Experience, it is necessary first to explain that, in this case, instead of having the full organization providing services to external customers, there is a department (Data Governance) that wants to provide a service (Data Lineage) to the other departments. However, instead of the Value Constellation Experience being for Data Governance, it is for Data Office, due to the collaboration that exists between the two departments inside Data Office (Data Governance and Data Quality).

It was therefore crucial to understand what the main activities of the Data Office are, and how the other departments could take advantage of what Data Governance is offering. It is crucial also to understand that the VCE is not a direct flow. In the case of Data Governance there is a set of activities that can happen at any time, and not all activities are performed in a single experience.

Based on the interviews, a set of activities were defined that the other departments could perform using the work and projects that are already made and the projects that are in the scope of Data Governance, understanding also that none of these activities are done without the help of external factors.

#### SERVICE EXPERIENCE

Based on the goals of the organization, the Service Experience is based on how the customer, in this case the other departments, can access and analyse the Metadata platform and take advantage of the work done by the Data Governance team. In other words, it shows every

service encounter between the Data Governance team and the customers, with the tool of Talend Data Catalog being the intermediary of these encounters. This is also the level where the most improvements were found, based on the different possibilities of the use of Data Lineage.

It is important to understand that in the current phase, there is no visibility and use of the tool in which the department works, so in a "before" phase, the Metadata Service Experience focuses on the work that is performed by the Data Governance team (in this case the Connection to Databases and Creation of Business Terms and Data Lineage). There are however two activities where the customer interacts with the team: having an Information Need and customer Access to Platform.

#### SERVICE ENCOUNTER EXPERIENCE

For the Service Encounter Experience, the focus is on the main activity inside the Metadata Service Experience: Access to Platform. This activity is where the main change happens, since, the main goal is for the other departments to start to use the platform (in this case Talend Data Catalog) with the implementation of Data Lineage. So, without having Data Lineage implemented, it was acknowledged that the main activities inside the Access to Platform were the login, the search of desired information and analysis of data. These activities will undergo a major change when Data Lineage is implemented.

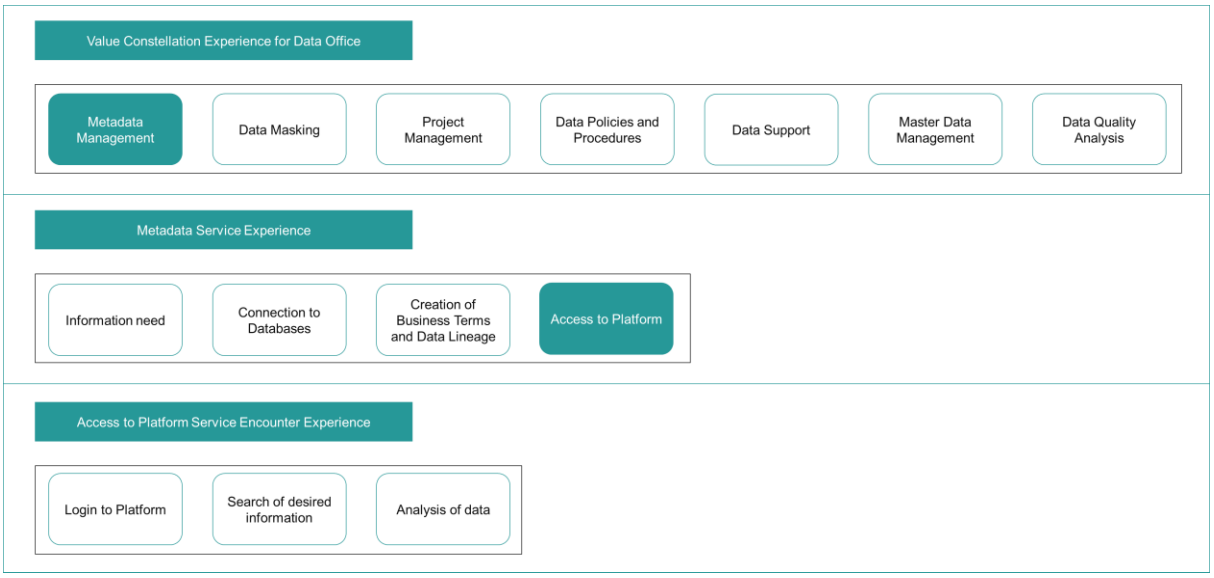


**Figure 7 Customer Experience levels**

### 5.1.3 Designing the Service Concept

#### CUSTOMER VALUE CONSTELLATION

As previously explained, the Customer Value Constellation represents the set of service offerings and interrelationships. In this case, the Customer Value Constellation is not going to suffer any modification, since the new service is going to be implemented in the Service System level.

In this case, due to the type of service, the phases represented in the Value Constellation Experience are the activities that are represented in the Customer Value Constellation.

Based on the Value Constellation Experience, it was noticed that the focus of the service offering was on the different activities that the Data Office (the combination of Data Governance and Data Quality) does on a daily basis, for example, working with Metadata Management or Master Data Management, supporting projects in different phases of the Project Management, or providing support to difficulties of different departments. However, it was noticed that the customer gets these services with the support of other external factors, for example, other Data departments (Data Services), or the Architecture department, due to the relationship between the two different areas. Other external factors are the infrastructure that is used and the Software Services and the respective Software Companies. So, this Customer Value Constellation focuses on the different activities of Data Office and the external factors.

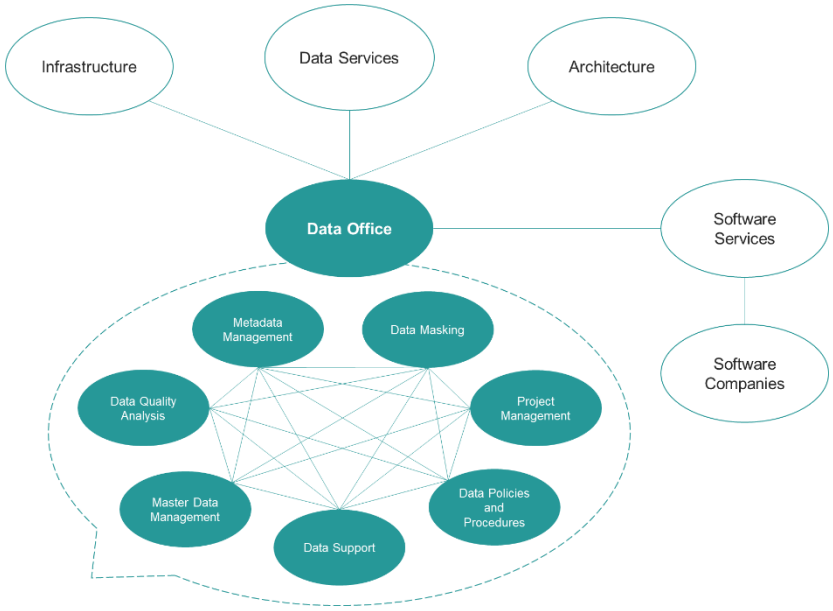


**Figure 8 Customer Value Constellation**

### 5.1.4 Designing the Service System

#### SERVICE SYSTEM ARCHITECTURE

Based on the Service Experience, it was possible to build the current SSA, with the activity of Metadata Management. As previously explained, in the Metadata Service Experience there are two main activities performed by Data Governance (Connection to Databases and Creation of Business Terms and Data Lineage) in order for the other departments to Access the Platform.

In the Service System Architecture there are four actors: the Customer, in this case, the other departments of Euronext, the Service Interface, divided in Talend Data Catalog (the main platform on creation and access of data) and the Data Governance Team and the Backend Support, being the backstage system of Talend Data Catalog.

In the old Service System Architecture (represented in figure 9), it is possible to see that the Customer has almost no interaction with the Data Governance team, because, without the implementation of Data Lineage, the platform had no visibility to the rest of the company. So, most of the activities were performed by Data Governance, directly or indirectly (connection to databases, import of the metadata, creation of business terms or data lineage mappings), based on the need previously communicated by the customer to the Data Governance team. The

Business Terms are terms created based on the definition of each field. The main goal of the Business Terms is to facilitate the comprehension of the fields, since, some may have difficult names or incomprehensible codes.

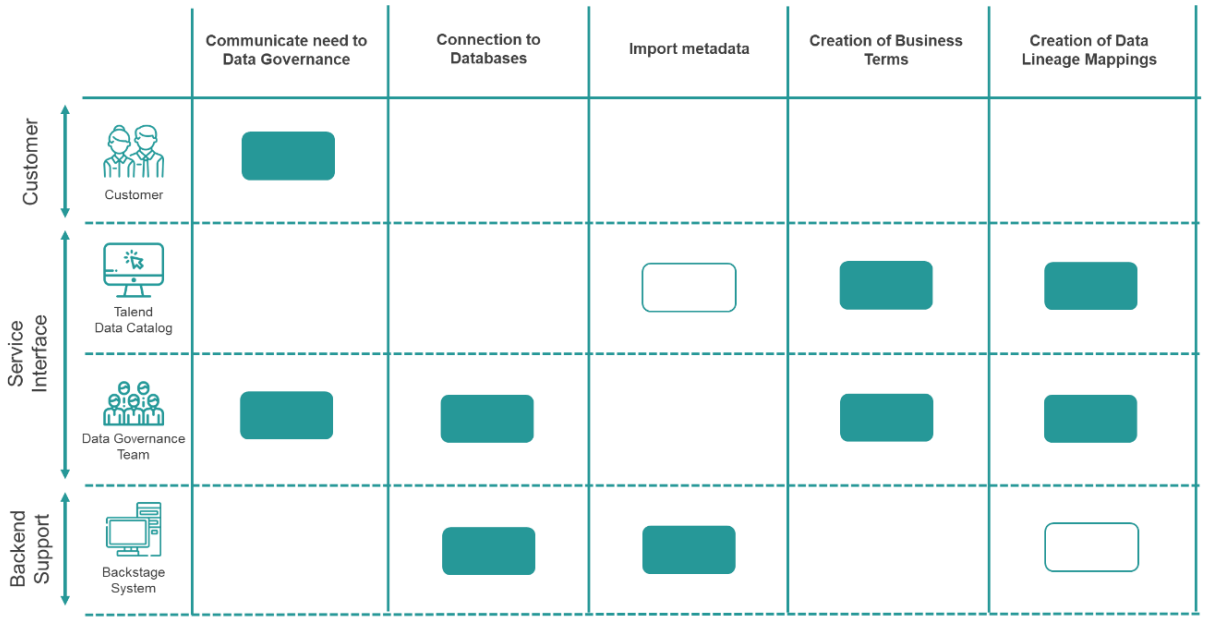


**Figure 9 Old Service System Architecture**

The main goal with the new Service System Architecture (represented in figure 10) is to introduce the analysis of Data Lineage and the experience of being able to perform Impact Analysis and to understand better the whole flow of data. With the new Service System Architecture, it is possible to see that the activities performed by the Data Governance team continue to be carried out.

The Analyse DB and Business Terms activities are together, since it is not a compulsory activity, and the user can do it or not. Next, a new activity of Search Data Lineage Mapping was added, so that now it can be a main activity when searching for information on Talend Data Catalog. The New Service System Architecture is composed by: Registration, Analyze DB's/Business Terms, Search Data Lineage Mapping, Report Difficulties, Perform Analysis and Decision Making.

With the Report Difficulties activity, the goal is to understand if the user has issues while navigating the interface, or if any information is missing or incorrect. This way, the Data Governance team can provide support. The last two activities, Perform Analysis and Decision Making are more related to the type of project that the user is working on, so these activities are more external to the Service Interface and the Data Governance team.

#### SERVICE SYSTEM NAVIGATION

The Service System Navigation (Figure 11) follows the same structure as the Service System Architecture. The main difference is that in SSN, the activities of the different actors are explained, and there is a flow that the customer follows since the beginning of the process until the last activity. So, since the beginning, the Customer needs to Communicate the need to the Data Governance team. Based on the need communicated to the Data Governance team, the

activities of connection to the databases and creation of the Business Terms and Data Lineage mapping may need to happen again. The next step depends on whether the person already has an account. The main point in this process is the Search Data Lineage Mapping, where the user can communicate a difficulty, like previously mentioned, asking for help to understand some information or report missing or incorrect information.

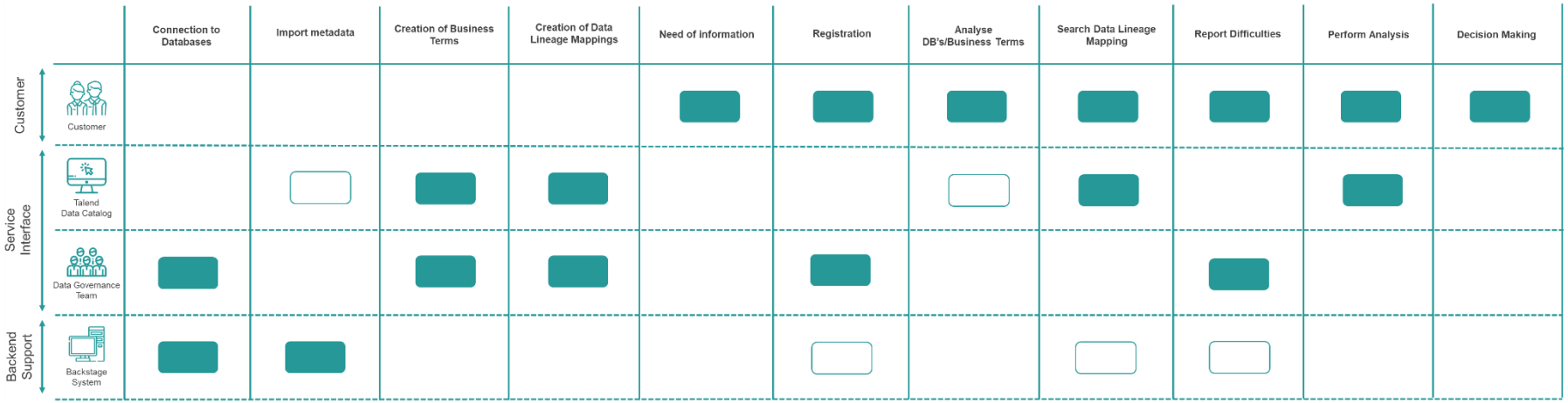


**Figure 10 New Service System Architecture**

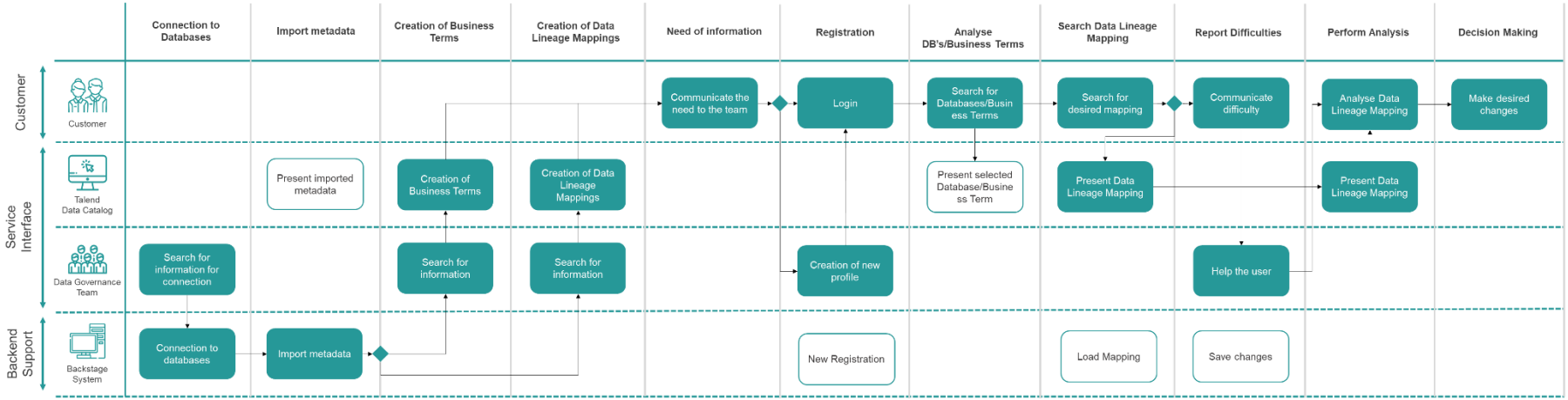


**Figure 11 Service System Navigation**

### 5.1.5 Designing the Service Encounter

#### SERVICE EXPERIENCE BLUEPRINT

The Service Experience Blueprint (Figure 12) is made based on the main interactions between the Customer, the Service Interface, and the Backstage Support. Besides these main interactions, the Service Experience Blueprint is differentiated from the Service System Architecture and Service System Navigation due to the possibility of understanding what the line of visibility and line of interaction is, the fail points (the F inside the red circle) and waiting points (the W inside the yellow circle) and the service interface links.

The Service Experience Blueprint was made for the activities “Search Data Lineage Mapping” and “Report difficulties”, since these are the two activities where the interaction between the Customer and the Service System is stronger.

Once the customer enters the Service Interface, it is possible to find the Data Lineage mappings in two ways: via tables on the Databases, in which the Service Interface presents the full Data Lineage and Data Flow, and Via Job ETL, in which the Service Interface presents each transformation individually. This means that the customer can choose the type of data to be analysed. Once the customer searches for the desired mapping, two things can happen: the mapping can be available, and this way, the customer can move on to analyse the data; or the mapping is not available, forcing the Data Governance team to search the information to do the Data Lineage, and this way the Customer can finally do the analysis. Depending on the type of customer (working on a more technical or business department), it can be noticed that some information or data may be wrong or missing, and the Data Governance team can be warned about the mistake. It is the job of the Data Governance team to understand if the information that came from the Customer is right or wrong, and the team can perform the changes or not. Having these full steps, the Customer can perform the Analysis Impact, which is an activity that is going to depend on the type of project that the Customer is currently working on.

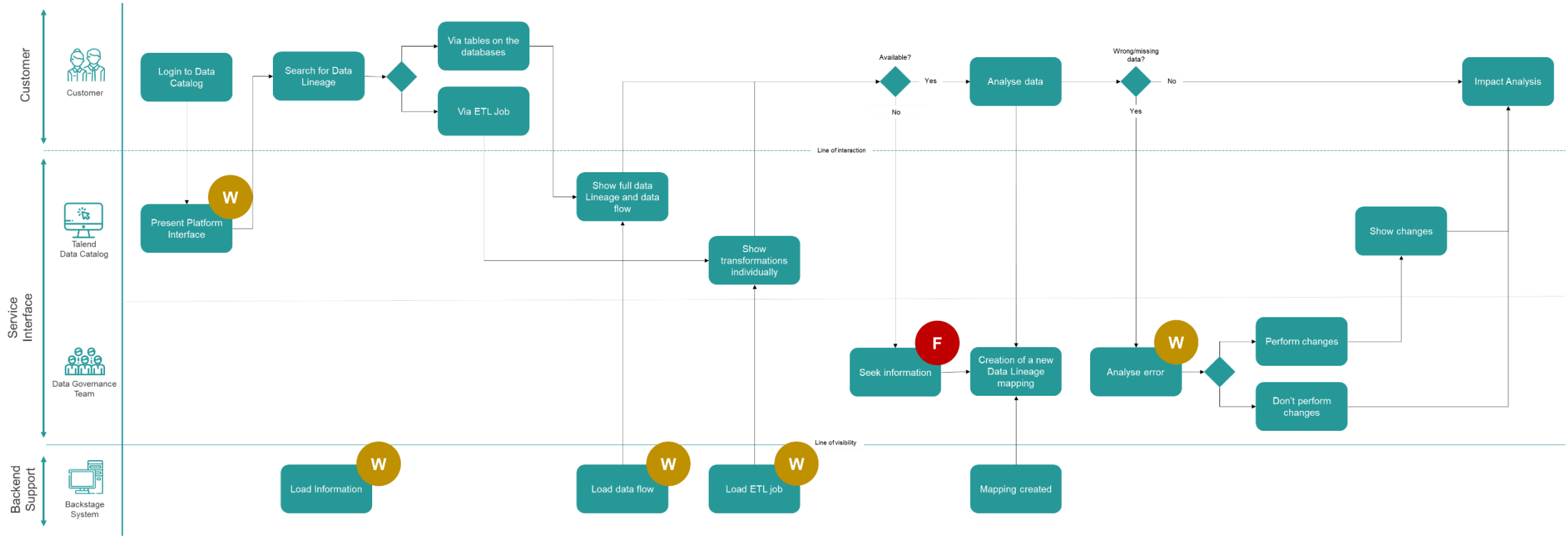


**Figure 12 Service Experience Blueprint**

## 5.2 Implementation and Validation

### 5.2.1 Implementation Steps

After the mapping of the Data Lineage and the design of the service through the different levels of Multilevel Service Design, it was necessary to develop a plan in order to implement the service inside the organization. This plan was created taking into consideration the needs of the department and the organization.

The implementation steps were created based on the conceptual model of Aarons, Hurlburt and Horwitz (2011): **Exploration, Adoption Decision/Preparation, Active Implementation, Sustainment.** Although this conceptual model was built for Implementation in Public Service Sectors, the different phases match the needs of this work. However, each phase will have a different approach than the authors had, applying them to Data Lineage.

With the help of the Data Governance team, a plan of implementation was designed:

1. **Exploration:** the exploration phase, being the "awareness of either an issue that needs attention or of an improve approach to an organizational challenge" (Aarons, Hurlburt and Horwitz, 2011), had its start in an earlier phase when the Data Governance team started to put Data Lineage in the Roadmap due to its benefits for the company;
2. **Adoption Decision/Preparation:** the adoption decision/preparation phase "is often conceptualized as a one-time event while, in practice, organizations may experiment with an innovation, sometimes intermittently, prior to broader implementation". This step happened when the decision to start this study was taken;
3. **Active Implementation:**
    a. **Gather information in a presentation**: the first step consists on gathering all the information and documentation produced in a presentation; this presentation is going to serve as the primary material in a presentation that needs to be done to the different departments:
        i. The first documentation is the analysis of the interviews that were presented previously in the section Results, that can show and explain with greater detail what Data Lineage is, without being too technical;
        ii. The presentation is also going to cover and explain a real project that is currently being done in the organization and uses Data Lineage as a main step to gather information.
    b. **Presentation of the material:** this presentation is crucial because one of the main goals of the Data Governance team is to put Data Lineage in the scope of all the departments. For that, this presentation needs to be done to five departments:
        i. **Change Management**: Every time there is a change in a Database, Schema or Table (specially removal or change of names), Data Governance needs to be warned, in order to prevent the Data Lineage mappings from turning obsolete;
        ii. **Project Management Office**: Data Lineage can be of great benefit if included in the Project Management Policy, within the Study Phase, due

to the ability to perform Impact Analysis in projects where data in general can have a great impact;

iii. **Problem Management**: If the type of problem is related to data or databases, Data Lineage can be a useful tool to analyse and fix the problem;

iv. **ABS:** it is important to maintain a certain level of communication with ABS (Application and Business Services), due to the documentation and knowledge present inside this department;

v. **Development Team:** The Development Team will be one of the main users of Data Lineage.

c. **Design Use Workflow for each department:** based on the individual needs, each department has a different way of using Data Lineage, and due that, based on the presentations that are made, another meeting with each department is going to be made, in the form of a brainstorming to design and develop the method of use of Data Lineage;

d. **Provide training sessions:** training sessions will be scheduled and available to every interested employee of Euronext, in order to have everyone on the same page when accessing, using, or analysing Data Lineage;

e. **Writing a Procedure and Policy**: as previously shown in the Customer Value Constellation of Data Governance, the writing of Procedures and Policies related to data is a main activity inside Data Governance, so, with the help of Multilevel Service Design (with the interviews and the Service Experience Blueprint) it is necessary to write the Procedure and Policy of Data Lineage.

4. **Sustainment:** The last phase, Sustainment, is used to "denote the continued use of an innovation in practice" (Aarons, Hurlburt and Horwitz, 2011). Although the implementation of Data Lineage is only at the Active Implementation phase, one of the goals is to continue to work in this subject and find new ways to improve the efficiency and efficacy of the creation and use of Data Lineage mappings.

A timesheet was designed and completed with the dates for each of the phases inside the Active Implementation step. However, due to scheduling problems, it was not possible to follow the dates and this step is still on-hold.

### 5.2.2 Validation Process

Being in an organizational context, it is necessary to have several validation phases throughout the whole project, and every modification needs to be validated by the teams involved in the work/project. In this case, the team of Data Governance, the main interested team, was involved in every decision and step that was taken:

- For the first step of the practical work, the interviews to the stakeholders, brainstorming sessions were made in order to understand what was the list of interviewees and the questions that should be addressed;
    - At the end of the phase of interviews, the results were discussed and analysed together, in order to design the presentation previously mentioned in the Implementation steps (1. Gather information in a presentation).
- The design of the service using Multilevel Service Design was validated by the team, to analyse the whole process and to understand the whole service;

- The implementation steps were brainstormed in a session with the Data Governance team and the Data Quality team, in order to have an external participant that would also have knowledge of implementation and the management phases of the projects.

## 6 Conclusion and future work

The goal of this research was to understand how to look at a framework like Data Lineage from a Service perspective, where Data Governance is the "service provider" and the "customers" are the employees of the different departments that are going to use the service. For that, a methodology of Service Design was used, Multilevel Service Design, from Fisk, Patrício and Constantine (2011), that divides the Service Design process in four main phases: (1) Understanding the Customer Experience, (2) Designing the Service Concept, (3) Designing the Service System and (4) Designing the Service Blueprint. This way, it is possible to design the service in full detail, focusing on the main levels of the service (concept, system, and blueprint).

However, before the use of Multilevel of Service Design, there was the need to expand the scope of Data Lineage, that is, using the steps explained in the sub-section Data Lineage mappings, inside the Problem Characterization section, several mappings were created in order to have a more cohesive work. This work is also shown in **Designing the Service System** from Multilevel Service Design, where the work of Data Governance is shown before the access of the customers to the platform.

During the use of Multilevel Service Design, the phase of **Understanding the Customer Experience** in Multilevel Service Design was used for two different goals: not only to understand the experience of the different departments when it comes to working with Data Governance, but also to understand and analyse how the main stakeholders perceive Data Lineage and what they think are the main benefits. This was an important step due to the information that was collected and used to present Data Lineage to other users that may not be much familiarized with Data Lineage.

The Service Blueprint designed in the last level of Multilevel Service Design, **Designing the Service Blueprint,** was used to create and write the Data Lineage procedure as a start-point.

It was interesting to look at Data Lineage from a different point of view as the ones analysed during the Literature Review, where there was a more technical and not as much focused on an organizational point of view.

To understand the full study, it is important to answer to the initial research questions and analyse the results:

1. **How can Data Lineage be implemented as an Internal Service, in order to mitigate regulatory risks and requirements?** It was important to understand what the main users or stakeholders think of and perceive Data Lineage in order to prove the value of Data Lineage. From the beginning when Data Governance set the goal to implement Data Lineage in the company, there was a clear view that this framework had several benefits for the rest of the company, and due to its characteristics, capabilities and benefits, it could fit under the "service" concept, building up a relationship of "service providers" and "service customers" with the other departments. What was missing was the full plan on how to implement it.
2. **Based on the needs of a company in the Financial sector, is Multilevel Service Design an appropriate service design methodology for such service?** Service Design is an area with several techniques, and in this case, the focus was on the type of service designed and not on the sector of the company. However, it was necessary to use a methodology that could deconstruct the service in different levels, making it possible to

understand the current activities that Data Governance can provide to the other departments, the semi-detailed explanation of the Metadata management, and an overview on how the customers can interact with the Data Lineage.

3. **What is the best way to implement such new service?** In this case, the whole Implementation phase started from the moment when Data Lineage started to be mapped on Talend Data Catalog. However, the plan was not clear, so, along with the Data Quality department, Data Governance set a plan (explained in the Implementation steps sub-section). The steps were planned taking into consideration the needs of the company and what was necessary to include the stakeholders in this process, so that the communication in fully clear.

The Service Design methodology helped implementing this service because Data Governance was able to:

- Analyse and explore in full detail Data Lineage;
- Brainstorm with different departments and stakeholders the capabilities of Data Lineage;
- Create a plan of implementation, with the information collected;
- Gain the attention of other actors that were not previously included in the project.

For Future Work, it is important to mention that the implementation steps planned for Data Lineage will continue to be followed, but due to time and scheduling problems, it is hard to predict when it will it finish. Once the implementation is finished, there will be continuous work related to Data Lineage, not only to expand the scope of the data flows, but also on ways how to improve the service offering and its value inside Euronext.

Other concepts related to Services can be also be studied, for example, Service Innovation, and how it can be related to Data Lineage. Service Innovation can be defined by three main components (Lusch and Nambisan, 2015):

- **Service Ecosystems**: as emergent A2A (actor-to-actor) structures actors create and recreate through their effectual actions and which offer an organizing logic for the actors to exchange service and cocreate value;
- **Service Platforms**: which enhance the efficiency and effectiveness of service exchange by liquefying resources and increasing resource density (facilitating easy access to appropriate resource bundles) and thereby serve as the venue for innovation;
- **Value Cocreation**: which views value as cocreated by the service offer(er) and the service beneficiary (e.g., customer) through resource integration and indicate the need for mechanisms to support the underlying roles and processes.

## References


Aarons, G. A., Hurlburt, M. and Horwitz, S. M. C. (2011) 'Advancing a conceptual model of evidence-based practice implementation in public service sectors', *Administration and Policy in Mental Health and Mental Health Services Research*, 38(1), pp. 4–23. doi: 10.1007/s10488-010-0327-7.

Alonso-Rasgado, M. T., Thompson, G. and Dannemark, O. J. (2005) 'State of the art in Service Design and Modelling', *VIVACE*, pp. 1–59. Available at: papers3://publication/uuid/4D92BF20-1727-465D-98CB-1D5B05DC716A.

Asmundson, I. (2012) *Financial Services: Getting the Goods, International Monetary Fund*. Available at: https://www.imf.org/external/pubs/ft/fandd/basics/finserv.htm.

Berg, B. L. and Lune, H. (2017) *Qualitative Research Methods for the Social Sciences*. 9th edn. Edited by Pearson. Boston.

Bose, R. (2002) 'A conceptual framework for composing and managing scientific data lineage', *Proceedings of the International Conference on Scientific and Statistical Database Management, SSDBM*, 2002-Janua, pp. 15–19. doi: 10.1109/SSDM.2002.1029701.

Bose, R. and Frew, J. (2005) 'Lineage retrieval for scientific data processing: A survey', *ACM Computing Surveys*, 37(1), pp. 1–28. doi: 10.1145/1057977.1057978.

Brown, P. and Stonebraker, M. (1995) 'BigSur: A System For the Management of Earth Science Data', *VLDB Conference*, pp. 720–728.

Buneman, P., Khanna, S. and Tan, W. C. (2001) 'Why and where: A characterization of data provenance?', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 316–330. doi: 10.1007/3-540-44503-x_20.

Buneman, P., Maier, D. and Widom, J. (2000) *Where was your data yesterday, and where will it go tomorrow? Data Annotation and Provenance for Scientific Applications*.

Business Dictionary (2020) *Compliance*. Available at: http://www.businessdictionary.com/definition/compliance.html.

Cadwallader, S. *et al.* (2010) 'Frontline employee motivation to participate in service innovation implementation', *Journal of the Academy of Marketing Science*, 38(2), pp. 219–239. doi: 10.1007/s11747-009-0151-3.

Chamberlain, P., Brown, C. H. and Saldana, L. (2011) 'Observational measure of implementation progress in community based settings: The Stages of implementation completion (SIC)', *Implementation Science*, 6(1), pp. 1–8. doi: 10.1186/1748-5908-6-116.

Chuang, P. T. (2007) 'Combining service blueprint and FMEA for service design', *Service Industries Journal*, 27(2), pp. 91–104. doi: 10.1080/02642060601122587.

Clarke, D. G. and Clark, D. M. (1995) 'Lineage', *Elements of Spatial Data Quality*, pp. 13–30.

Crews, K. R. *et al.* (2011) 'Development and implementation of a pharmacist-managed clinical pharmacogenetics service', *American Journal of Health-System Pharmacy*, 68(2), pp. 143–150. doi: 10.2146/ajhp100113.

Cui, Y., Widom, J. and Wiener, J. L. (2000) *Tracing the lineage of view data in a warehousing environment, ACM Transactions on Database Systems*. doi: 10.1145/357775.357777.

DATAVERSITY (2017) *Data Lineage Demystified: The What, Why, and How*. Available at: https://www.dataversity.net/data-lineage-demystified/.

DATAVERSITY (2020) *About DATAVERSITY - Mission Statement, Events, & Founder*. Available at: https://www.dataversity.net/about-dataversity-net/.

Eagan, P. D. and Ventura, S. J. (1993) 'Enhancing value of environmental data: Data lineage reporting', *Journal of Environmental Engineering (United States)*, 119(1), pp. 5–16. doi: 10.1061/(ASCE)0733-9372(1993)119:1(5).

Euronext (2020) *Euronext Homepage*. Available at: https://www.euronext.com/pt.

Fixsen, D. L. *et al.* (2009) 'Core implementation components', *Research on Social Work Practice*, 19(5), pp. 531–540. doi: 10.1177/1049731509335549.

Frew, J. and Bose, R. (2001) 'Earth system science workbench: A data management infrastructure for earth science products', *Proceedings of the International Conference on Scientific and Statistical Database Management, SSDBM*, pp. 180–189. doi: 10.1109/ssdm.2001.938550.

Glavic, B. and Dittrich, K. (2007) 'Data provenance: A categorization of existing approaches', *Datenbanksysteme in Business, Technologie und Web, BTW 2007 - 12th Fachtagung des GI-Fachbereichs 'Datenbanken und Informationssysteme' (DBIS), Proceedings*, (June), pp. 227–241. doi: 10.5167/uzh-24450.

Grenha Teixeira, J., Pinho, N. F. de and Patrício, L. (2019) 'Bringing service design to the development of health information systems: The case of the Portuguese national electronic health record', *International Journal of Medical Informatics*. Elsevier, 132, p. 103942. doi: 10.1016/j.ijmedinf.2019.08.002.

Hevner, A. R. *et al.* (2004) 'Design science in information systems research', *MIS Quarterly: Management Information Systems*, 28(1), pp. 75–105. doi: 10.2307/25148625.

Hoang, D. T. A. (2011) 'Impact analysis for on-demand data warehousing evolution', *CEUR Workshop Proceedings*, 789, pp. 280–285.

Holmlid, S. and Evenson, S. (2008) 'Bringing Service Design to Service Sciences, Management and Engineering', in *Service Science, Management & Engineering (SSME)*, pp. 341–345. doi: 10.1007/978-0-387-76578-5_50.

Karpen, I. O., Gemser, G. and Calabretta, G. (2017) 'A multilevel consideration of service design conditions', *Journal of Service Theory and Practice*, 27(2), pp. 384–407. doi: 10.1108/jstp-05-2015-0121.

Karwan, K. R. and Markland, R. E. (2006) 'Integrating service design principles and information technology to improve delivery and productivity in public sector operations: The case of the South Carolina DMV', *Journal of Operations Management*, 24(4 SPEC. ISS.), pp. 347–362. doi: 10.1016/j.jom.2005.06.003.

Khatri, V. and Brown, C. V. (2010) 'Designing data governance', *Communications of the ACM*, 53(1), pp. 148–152. doi: 10.1145/1629175.1629210.

Kim, M. J. *et al.* (2018) 'Approach to service design based on customer behavior data: a case study on eco-driving service design using bus drivers' behavior data', *Service Business*. Springer Berlin Heidelberg, 12(1), pp. 203–227. doi: 10.1007/s11628-017-0343-8.

Korper, A. K. *et al.* (2020) 'Service design as an innovation approach in technology startups: a longitudinal multiple case study', *Creativity and Innovation Management*, 29(2), pp. 303–323. doi: 10.1111/caim.12383.

Lanter, D. P. (1991) 'Design of a Lineage-Based Meta-Data Base for GIS', *Cartography and Geographic Information Systems*, 18(4), pp. 255–261.

Lovelock, C. and Wirtz, J. (2011) *Services Marketing: People, Technology, Strategy*. 7th edn. Edited by N.J. Upper Saddle River: Pearson/Prentice Hall.

Lusch, R. F. and Nambisan, S. (2015) 'Service innovation: A service-dominant logic perspective', *MIS Quarterly: Management Information Systems*, 39(1), pp. 155–175. doi: 10.25300/MISQ/2015/39.1.07.

McCombes, S. (2019) *Developing Strong Research Questions | Criteria and Examples*, *Scribbr*. Available at: https://www.scribbr.com/research-process/research-questions/.

Michel, S., Brown, S. W. and Gallan, A. S. (2008) 'Service-logic innovations: How to innovate customers, not products', *California Management Review*, 50(3), pp. 49–65. doi: 10.2307/41166445.

Moeller, S. (2008) ‘Customer integration - A key to an implementation perspective of service provision’, *Journal of Service Research*, 11(2), pp. 197–210. doi: 10.1177/1094670508324677.

Morelli, N. (2009) ‘Service as value co-production: Reframing the service design process’, *Journal of Manufacturing Technology Management*, 20(5), pp. 568–590. doi: 10.1108/17410380910960993.

Mwebaze, J. *et al.* (2010) ‘A data lineage model for distributed sub-image processing’, in *ACM International Conference Proceeding Series*, pp. 209–219. doi: 10.1145/1899503.1899527.

Nilsen, P. (2015) ‘Making sense of implementation theories, models and frameworks’, *Implementation Science*, 10(1), pp. 1–13. doi: 10.1186/s13012-015-0242-0.

Patrício, L. *et al.* (2011) ‘Multilevel service design: From customer value constellation to service experience blueprinting’, *Journal of Service Research*, 14(2), pp. 180–200. doi: 10.1177/1094670511401901.

Ready.gov (2015) *Business Impact Analysis*. Available at: https://www.ready.gov/business-impact-analysis.

Rossmann, D. and Young, S. W. H. (2020) ‘Evidence based practice for virtual reality spaces and services: A service design case study’, *Evidence Based Library and Information Practice*, 15(2), pp. 143–149. doi: 10.18438/eblip29714.

Search Data Management (2020) *What is data governance and why does it matter?* Available at: https://searchdatamanagement.techtarget.com/definition/data-governance.

Shostack, G. L. (1984) ‘Designing services that deliver (Blueprint)’, *Harvard Business Review*, (84115), pp. 132–139.

Simmham, Y. L. (2015) ‘A Survey of Data Provenance in e-Science’, *SIGMOD*, 34, pp. 31–36.

Spery, L., Claramunt, C. and Libourel, T. (2008) ‘A lineage metadata model for the temporal management of a cadastre application’, in *Proceedings of the 10th International Workshop on Database & Expert Systems Applications (DEXA ’99)*, pp. 466–474. doi: 10.1109/dexa.1999.795211.

Spillane, J. P., Reiser, B. J. and Reimer, T. (2002) ‘Policy implementation and cognition: Reframing and refocusing implementation research’, *Review of Educational Research*, 72(3), pp. 387–431. doi: 10.3102/00346543072003387.

Subramanian, G. *et al.* (2019) ‘Data Lineage Identification and Change Impact Prediction in a Distributed Computing Environment’. United States.

Talend (2020) *Talend - A Cloud Data Integration Leader*. Available at: https://www.talend.com/.

Thomas, W. *et al.* (2009) ‘Exploring Service Blueprints for Multiple Actors: A Case Study of Car Parking Services’, *First Nordic Conference on Service Design and Service Innovation*, pp. 1–11.

Vargo, S. L. and Lusch, R. F. (2004) ‘Evolving to a New Dominant Logic for Marketing’, *Journal of Marketing*, 68(1), pp. 1–17. doi: 10.1509/jmkg.68.1.1.24036.

Vargo, S. L. and Lusch, R. F. (2016) ‘Institutions and axioms: an extension and update of service-dominant logic’, *Journal of the Academy of Marketing Science*, 44(1), pp. 5–23. doi: 10.1007/s11747-015-0456-3.

Wang, Y. H., Lee, C. H. and Trappey, A. J. C. (2017) ‘Service design blueprint approach incorporating TRIZ and service QFD for a meal ordering system: A case study’, *Computers and Industrial Engineering*. Elsevier Ltd, 107, pp. 388–400. doi: 10.1016/j.cie.2017.01.013.

Warren, K. (2020) *Qualitative vs. Quantitative Research*, *GradCoach*. doi: 10.5840/cpsem198946.

Woodruff, A. and Stonebraker, M. (1997) ‘Supporting fine-grained data lineage in a database visualization environment’, in *Proceedings - International Conference on Data Engineering*,

pp. 91–102. doi: 10.1109/icde.1997.581742.

Zeithaml, V. A., Parasuraman, A. and Berry, L. L. (1985) ‘Problems and Strategies in Services Marketing’, *Journal of Marketing*, 49(2), p. 33. doi: 10.2307/1251563.