Dandi Merga Gutema

Data Governance Practices in Higher Education Institutions:

Opportunities and Challenges

ACKNOWLEDGEMENT

First, I would like to express my deepest gratitude to my supervisors Professor Jose Teixeira and Professor Kristina Eriksson-Backa for their consistent guidance throughout the process of writing this thesis. Without their persistent advice and support, this thesis could not have come into an end.

Second, I would also like to thank all participants of the interviews who gladly gave me their valuable time during data collection process. Without their passionate involvement and cooperation, the interview data could not have been successfully gathered.

Third, I would like to express my heartfelt thanks to my family and friends for their endless encouragement and support. Without their care and support, this accomplishment would not have been possible.

Finally, I am also grateful to all the staff of Åbo Akademi University for their excellent assistance and advice.

ABSTRACT

ÅBO AKADEMI UNIVERSITY – Faculty of Social Sciences, Business and Economics, and Law (FSEJ)

Subject: Governance of Digitalization
Writer: Dandi Gutema
Title: Data Governance Practices in Higher Education Institutions: Opportunities and Challenges
Supervisor: Jose Teixeira   Supervisor: Kristina Eriksson-Backa

Higher Education Institutions (HEIs) collect data from their internal operations, external databases, and research activities on a daily basis. Previous studies indicated most HEIs were facing difficulties to unlock (i.e., to obtain) value from their data assets. Effective data governance can enhance HEIs to manage their institutional data by establishing frameworks, responsibilities, accountabilities, policies, procedures, guidelines, and processes. This study attempted to bridge the gap by exploring the challenges and opportunities pertaining to data governance practices in Finnish Higher education Institutions (HEIs). An appropriate data governance framework was drawn from the literature which focuses on decision domains of data governance to meet the objectives of the study.

This study is qualitative research in which exploratory research approach was adopted to gain in-depth understanding of the research domain. Alongside, data was collected through semi-structured interviews. Heterogeneous purposive sampling was employed to choose experts from different roles. Professionals who were working ICT, Library, and administration areas at two HEIs were interviewed. The respondents were selected purposively based on their familiarity with the topic of this study and acquaintance with utilization of data for decision making. Moreover, qualitative research methods such as thematic analysis and deductive coding were used to analyze, and categorize the themes based on the chosen framework

The result of the study revealed that in the majority of the data governance decision domains, the HEIs have developed policies, procedures, and guidelines to ensure management of data assets and maintain quality of data. Moreover, by considering data as institutional assets, the HEIs were able to generate reports and the statistics, open

and reuse research data, and publications, comply with regulations and laws, measure academic performances, manage data security risks, and control data flow. The finding of the study also revealed challenges pertaining to data governance practices. Most of the challenges stemmed from the absence of clear policy, lack of expertise, lack of establishing comprehensive roles regarding data-related activities, and lack of plans in data quality audits and data inventory areas. The findings of the study showed that majority of data governance challenges are more strategic than operational. Therefore, decision makers at HEIs are required to bridge the identified gaps through establishing comprehensive framework, policies, procedures, and plans that covers all data governance decision domains.

Keywords: data governance, data management, higher education institutions, data quality, IT governance

Date: 10.05.2023	Number of pages: 100

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

BIBFRAME: Bibliographic Framework

DAMA: Data Management Association

DGI: Data Governance Institute

DMBOK: Data Management Body of Knowledge

DMP: Data Management Plan

DOI: Digital Object Identifier

EU: European Union

EU/EEA: European Economic Area

FAIR: Findable, Accessible, Interoperable, Re-Usable

GDP: Gross Domestic Product

GDPR: General Data Protection Regulation

HEI:Higher Education Institution

HEIs:Higher Education Institutions

IBM: International Business Machines

ICT: Information and Communication Technology

IPR: Intellectual Property Rights)

KMPG: Klynveld Peat Marwick Goerdeler

KPI: Key Performance Indicators

MARC: MAchine-Readable Cataloging

MFA: Multi Factor Authentication

Non-EU: Non- European Union

OECD: Organisation For Economic Co-Operation And Development

PDSA: Plan, Do, Study, Act

UAS: University Of Applied Sciences

USD: United States Dollar

US: United States

1 INTRODUCTION

1.1 Background of the study

Technology is crucial to the advancement of higher education (Reeves & Pearlman, 2017). According to Okoro (2021), organizations in both public and private sectors can benefit greatly from data governance. Companies can establish corporate-wide accountability for data, compliance with international privacy laws, and data transparency in their procedures concerning handling of personal data by using a data governance framework. Businesses will benefit from a data governance framework that will help them organize and record their data quality accountabilities.

The OECD (2022) report on education states that significant amounts of financial resources from national output were invested in education throughout OECD countries. The average percentage of gross domestic product (GDP) that OECD countries spent on education in 2019, from primary to tertiary educational institutions, was 4.9%, according to the report. In contrast, Finland spent 5.2% of its gross domestic product (GDP). Finland saw a 4% increase in overall funding for educational institutions between 2008 and 2019. Regarding tertiary-level education, Finland's expenditure per student is higher than that of other educational levels. The average annual cost per student in Finland is 18,129 USD. In Finland, research, and development (R&D) spending accounts for 47% of tertiary education spending, compared to 29% on average across OECD nations.

As stated by Khairi (2019), universities, like any other large enterprise, produce a vast amount of data that need to be managed and secured. Data governance policies raise more basic issues as data become more vital to academic practices (Madison, 2020). Koltay (2016) claims that data governance should not be optional for organizations because it increases transparency in data-related processes and reduces costs associated with managing of organizational data assets. In addition, it promotes organizational success by ensuring reliable compliance practices in accordance with existing regulatory frameworks and guidelines. Also, data governance helps organizations to establish data-related decision rights (i.e., accountabilities and responsibilities) by outlining decision domains. As data are everywhere, data governance operates horizontally (Koltay, 2016).

For the reasons mentioned above, it is necessary to study the present data governance practices in higher education institutions. The main aim of this study is to explore the challenges and opportunities pertaining to data governance practices in Finnish Higher Education Institutions (HEIs).

1.2 Statement of the problem

The majority of data governance measures in Finnish higher education, from the researcher's perspective, are implemented through the universities' data and information management plans, laws and regulations, and guidelines related to data processing. As a result, it is essential to consolidate the disparate data-related activities into a single framework, such as data governance.

The study conducted by Nielsen (2017) indicates that most papers published on data governance from 2007 to 2017, are from the areas of computer science and information systems with 34% and 32% of the total publications, respectively. As the same authors mentioned, in contrast, publications on data governance pertaining to education sector accounted for fewer than 5% of the total research papers. The researcher also revealed that most studies on data governance in the education discipline focus on the theoretical aspect. Therefore, in previous studies, the practical perspective of data governance has not been addressed properly.

Recent research on data governance in Finland have focused on different areas such as industrial machinery (Pennanen, 2014), forestry (Rantala et al., 2022), banking (Paananen, 2020) health (Dahlberg et al., 2017), and laws and regulations on data governance such as the Data Governance Act (Ruohonen & Mickelsson, 2023), among others. However, in education sector, data governance practices and implementation have not been studied extensively. By exploring the opportunities and constraints of data governance practices in Finnish Higher Education Institutions, this study seeks to close the existing gap.

Based on the research problems mentioned above, the following research questions were formulated.

RQ1: How are data governance practices carried out at higher education institutions?

RQ2: Which benefits do higher education institutions obtain from utilizing data governance practices?

RQ3: Which obstacles, both existing and potential, prevent data governance practices in higher education institutions?

1.3 Objective of the study
1.3.1 General objective of the study

The general objective of this study is to assess challenges and opportunities pertaining to data governance practices in Finnish higher education institutions.

1.3.2 Specific objectives of the study

To meet the general objective of this study mentioned above, the following specific objectives were formulated.

- ❖ To assess data governance practices at higher education institutions (HEIs).
- ❖ To select appropriate data governance framework to understand data governance decision domains.
- ❖ To employ the selected data governance framework to assess the opportunities and challenges pertaining to data governance practices at selected HEIs.
- ❖ To examine the opportunities or benefits of data governance practices at HEIs.
- ❖ To examine the challenges that hinder data governance practices at HEIs.

1.4 Significance of the study

This study attempted to explore the challenges and opportunities of data governance practices at higher education institutions in Finland. The researchers assumes that the following benefits can be gained from the findings of this study: (1) the results (assuming there are more than one) of the study may inspire others to conduct additional research in the area, (2) the findings (assuming there are more than one) of the study can help the selected higher education institutions to assess the status of the existing data governance practices of their institution. Therefore, the HEIs can utilize the findings of this study to initiate data governance program in the future, (3) other Finnish HEIs can use the findings

of this study to assess the challenges and opportunities of data governance practices at their institutions. The results of this study may also motivate Finnish HEIs to initiate institution-wide projects pertaining to data governance or to conduct in-depth research, (4) the HEIs selected for this study can assess their current policies, strategies, guidelines, and initiatives using the findings of the study. Additionally, they may use the results to create a detailed guideline or checklist that can assist successful implementation of data governance initiatives, (5) this study can also provide information for future researchers to comprehend the potential and problems relating to data governance practices at HEIs.

1.5 Scope and limitations

This study focuses on data governance practices in the educational sector. As mentioned above, there are fewer studies pertaining to this area. For this study, HEIs that share resources such as library resources, staff, courses, and education materials were selected purposefully. It is reasonable to consider that educational institutions share data among themselves regarding student grades, courses, teaching material, and electronic resources. Therefore, in this study, institutions that have been collaborating and sharing data with each other were selected because it is necessary to study data sharing and data access domains between HEIs. In addition, the locations of HEIs and subject matter knowledge were also considered when choosing the institutions. Accordingly, interview invitations were sent out to professionals from three different higher education institutions. However, eight experts (out of 15 interview invitations) from two HEIs accepted the interview invitations. Moreover, one interview was used for pilot testing whereas the remainders were considered for further analysis. Therefore, the sample size of this study reduced to seven specialists from two HEIs. Moreover, in this study, experts from the HEIs ICT department, administration, and library were only accounted for the interview process. Due to time limitation and availability issue professionals from data protection office, and finance department at HEIs were not included in the study sample. Therefore, further studies could consider involving those experts for detailed and comprehensive insights.

1.6 Structure of thesis

This study is organized into six chapters. The first chapter introduces the background of the study, statement of the problem, objectives, significance, limitations, and scope of the

study. The second chapter provides literature review. It consists of brief definitions of data, data governance concepts, data governance models, etc. It also describes the conceptual framework of the study. The third chapter is methodology. In this chapter, the whole process of the research design is described in detail. It consists of research purpose, data collection method, data types and sources, data analysis, interpretation, presentation methods, and data analysis tools. The fourth chapter presents the result of the study. The fifth chapter presents the discussion and conclusion. Lastly, chapter six provides recommendations for further research.

2 LITERATURE REVIEW

This chapter presents review of literature on the concepts of data governance, data management, IT governance, data governance frameworks, importance, and challenges of data governance application in Higher Education Institutions context. It also outlines the summary of data governance frameworks, and conceptual framework of the study.

2.1 What is data?

Many scholars defined data from different perspectives. According to DAMA (2017), from technological perspective, data can be defined as *"information that has been stored in digital form (though data is not limited to information that has been digitized)"*. As stated by Hoffer et al. (2016), from historical perspective data referred to facts pertaining to objects or specific events that could be captured and stored on computer media. The traditional definition of data supports structure of data and types of data. For example, customer name, address, phone number etc. However, in modern times, data represent multiple arrays of facts and objects such as emails, tweets, maps, location information, sounds, and videos among others (Hoffer et al. 2016). Thus, the nature and type have changed over years. Nowadays unstructured data, such as videos and audio, also called multimedia data, and structured data such as specific object name, can be combined to create real multimedia environment (Hoffer et al., 2016). According to the same author, a new definition that comprises both unstructured and structured data type is required. Thus, authors define data as *"a stored representation of objects and events that have meaning and importance in the user's environment"* (p.41). Therefore, these data (i.e., captured data) can be processed to drive profit (DAMA, 2017).

As argued by DAMA (2017) most people assume quite wrongly; that data as form truth that represents facts. With this, DAMA indicated that data itself is a representation that requires a context to interpret it. According to Smidt (2021), data are the most valuable property of the 21st century that help organizations to increase competitiveness, improve processes and drive business knowledge.

2.2 Data governance: definition and concepts

Before we go in detail to the definition and concept of data governance, let us consider first the definition of governance since the definition of data is already mentioned above. According to DAMA (2017), governance is described as " *the exercise of authority and control over a process, organization or geopolitical area".* Governance in Information Technology (IT) refers to who makes decisions about the effective management and use of IT resources and who is accountable for those decisions (Khatri & Brown, 2010).

By combining the definitions of both data and governance, DAMA (2017) defines data governance *"as the exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets."* Smidt (2021) states that organizations are struggling to acquire values from their data assets they store and process. Thus, data governance can solve the institutional problems by providing frameworks, guidelines, policies, and processes which helps organizations to utilize and manage organizational data assets (Smidt, 2021). According to Ladley (2012,), data governance can also be defined as *"the organization and implementation of policies, procedures, structure, roles, and responsibilities which outline and enforce rules of engagement, decision rights, and accountabilities for the effective management of information assets"* (p.11). In general, the aim of data governance is to increase the value organizational data asset (Otto, 2011). Figure 1 depicts fundamental concepts in data governance:
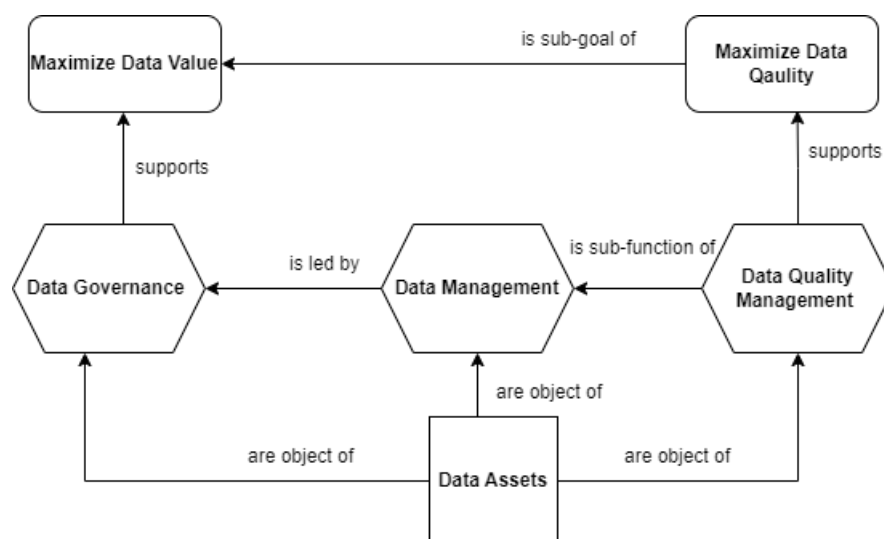
Figure 1 Fundamental concepts in data governance (Otto, 2011)

2.3 Data management and Data governance

Previous studies indicate that there is fundamental difference between data management and data governance. As shown in Figure 1, data management focuses on maximizing the quality of data whereas data governance focuses on enhancing the value of organizational data assets (Otto, 2011).

According to DAMA (2017), data management is defined as *"development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles"*. Data management calls for both technical and business skills to make reliable decision pertaining to maximizing the quality and value of data and deploy databases (DAMA, 2017). Alongside the maximization of data quality and uses, data management assists organization to integrate their data and knowledge resources which in turn boost organizational knowledge discovery and innovation initiatives (Wilkinson et al., 2016). The outputs of such initiatives can also be reused for further innovation practices.

Generally, the aim of data management is to make sure that the organizations gain the value of their data asset while data governance concerns with how data decisions are made and how people and processes are expected to behave in relation to data (DAMA, 2017). To sum up, Koltay (2016) points out that data management involves determining the actual criteria employed for data quality, whereas data governance is about designating who should make these decisions.

2.4 IT governance Vs. Data governance

As stated by DAMA (2017), data governance is different from IT governance. The same author concerns with decision making regarding IT investment, IT project portfolio, and IT application portfolio by aligning IT strategies with organizational goals whereas data governance solely empathizes on management of institutional data assets. According to Khatri & Brown (2010), IT governance concerns with identifying who holds decision making rights and who is accountable for making decision pertaining to IT assets while data governance concerns with identifying the authority who is accountable and responsible to make decision pertaining to the data asset of the organization. IT governance is directed toward designing a framework that defines decision rights and

accountability with respect to IT assets use (Weill & Ross, 2004). In short, IT governance concerns with utilization of IT assets of the organization while data governance focuses on governing data assets of organizations (Khatri & Brown, 2010; Weill & Ross, 2004).

2.5 Importance of data governance

Several definitions of data governance demonstrate its importance in diverse organizations. Newman and Logan (2006) point out that data governance is a collection of actions, principles, methods, norms, regulations, and technologies that are required to control, store, and leverage information as a business resource. In addition, another aspect of data governance describes data governance as the process by which an organization oversees the quantity, regularity, accessibility, safety, and accessibility of data (Mahanti, 2021). As mentioned by Thomas (2006), data governance applies to the organizational structures, guidelines, authority to make decisions, and responsibilities of people and information systems implementing data-related activities.

As mentioned by Cheong (2007), data governance plays an important role since it specifies guidelines and processes to ensure active and successful data management. Implementing a system for data governance provides opportunities for interaction across different levels of an organization when handling company data, plus the capacity to match multiple data activities with business goals (Cheong, 2007). As developments in IT enable organizations to gather information in both structured and unstructured formats, data governance has become increasingly crucial to many commercial and non-profit organizations. Cheong's (2007) study paper showed rising challenges linked to data management and data quality. In that study, he pointed out that the quality of the data reflects the decision maker's confidence in the data (Cheong & Chang, 2007). Data governance also specifies accountability for decision-making. It identifies who has the power to make decisions and who is held accountable for data property (Khatri & Brown, 2010).

Data Governance users in academic contexts include scholars, universities, financiers, media outlets, and people in general. Frameworks for governance are also required for managing social data because protecting sensitive information involves establishing rules

and norms of practice and encouraging behavioural shifts toward more powerful data governance (Hartter et al., 2013).

The necessity for data governance in higher education is not extremely popular, but there is a surge of pressure to modernize obsolete technology systems. Data governance is a necessary foundation that enables higher education institutions to provide reliable information on what they do, take decisions, and maintain data for the next generations. The American Council on Education have selected a research group where they have been employing data analytics for student success and sustainability to evaluate the possibility of student growth or support. It is helping them improve graduation rates and eliminate inequality more than before (Tolley & Shulruf, 2009). In recent years, educational institutions such as the University of Texas have formed innovative data-sharing collaborations with government departments, developing student performance and employment prospects (Horsman, 2018). They have used data to explore strategies to improve inclusion, equality, and diversity. According to the report, Georgia State University has raised graduation rates by more than 20% and closed equity disparities for economically disadvantaged pupils from minority backgrounds primarily due to work of establishing analytics platforms that began in the early 2000s.

Therefore, data governance establishes internal guidelines and data regulations for data collection, storage, retrieval, processing, and disposal. Therefore, it governs who can access data, what types of data, and what types are governed. On their digital transformation journeys, every firm requires data governance. It has become their most precious asset in the competitive market. Data governance cannot be considered an alternative because it helps company performance through consistent and legal procedures (Koltay, 2016).

2.6 Data governance challenges in education sector

Many studies provided helpful recommendations regarding ways to prepare and design data governance in different areas of business, but there are only a few studies that define and explain why it is challenging for higher education institutions to implement data governance into an evolving reality (Brous et al., 2016).

So far, the definitions and importance of data governance were addressed, but there are certain difficulties that can arise while handling data in an educational organization. Some researchers, for example, stated that the lack of uniform information throughout the university makes reporting on institution-wide operations challenging (Koltay, 2016). This causes a surplus of data and increases stress among managers, staff members, and academic divisions. Also, many organizations collect data from multiple units using different methods. They think they have gained insight from one report or another in this manner.

However, data varies from place to place, and as well as collection goals can be different, and it likewise varies differently amongst administrations and departments. As an outcome, gaining conclusions from a broader perspective becomes difficult (Al-Ruithe et al., 2018). This leads to operational uncertainty, in which multiple interpretations of the same data produce a lack of credibility and trust in the data and the persons in charge.

It is also often observed that educational institutions struggle to obtain an overview of the data that is available. Although there is an interest to construct suitable information facilities, the current design is very dispersed and plagued with old systems. Exploring potential value generation with data governance becomes more difficult when there is no visibility (Benfeldt et al., 2020). Developing data governance standards can be challenging without an in-depth knowledge of the ways data assets must contribute to organizational goals.

According to DAMA (2017), the essence of data governance program is to maintain the quality of institutional data assets. The quality of data can be affected by various factors such as absence of standards, lack of planning, incompatible system development, data silos, inadequate documentations, and lack of understanding the existing data assets (DAMA, 2017).

Another challenge of data governance is data breaches. Inaccurate reporting inside departments can cause problems, but false disclosure outside the organization can endanger the institution. It has the potential to permanently damage an institution's credibility, as well as lead to compliance violations and legal penalties. Previous associate vice president of information systems at the University of Texas Southwestern and the

University of Chicago stated that after they executed a data governance plan and gathered the data in one database, they excluded Health Insurance Portability and Accountability Act data and the Family Educational Rights and Privacy Act from the analysis for avoiding unnecessary risk (Horsman, 2018).

Data governance challenges should be viewed as a social action problem instead of a process of allocating responsibility. It can offer a completely distinct path for progress for educational institutions if the institutions can overcome these challenges (Alhassan et al., 2018). From the above discussion, we reason that the challenges concerning data governance are individuals in group effort that tend to give various meanings to the shared actions. Developing skills for data governance is challenging as it involves proficiently handling several assets provided by various parties toward a common goal accomplished jointly (Ansell & Torfing, 2016). Allowing communication among departments regarding data governance is also difficult since users prefer to behave in ways that support their priorities at the price of obtaining a stronger shared purpose.

2.7 Higher education in Finland context

The education system of Finland comprises multiple actors. These are early childhood education and care, pre-primary, basic and general upper secondary education, vocational education and training and higher education[1]. Higher education in Finland is divided in two broad categories, namely Universities and Universities of Applied Sciences (UAS). Universities are aimed at providing educations aligned with scientific research. They can award from bachelor's degree to doctoral degrees. In contrast, Universities of Applied Sciences (UAS) provide pragmatic or practice-based education that involves applied research and development to fill the gap of job market. Currently there are 14 universities and 24 Universities of Applied Sciences (UAS)[2]. Those universities are governed by the University act and UAS act respectively.

Both Universities and UAS in Finland also offer open studies to promote lifelong learning and continuous education based on their own syllabuses. According to Publications of the Finnish Government (2021), due to the increasing demand of "upskilling", Finnish

---

[1] https://www.oph.fi/en/education-system
[2] https://www.fulbright.fi/studies-and-research-finland/higher-education-finland

Higher Education Institutions, Universities and UAS, were attempted to facilitate continuous learning that could respond to the needs of the society. The report suggests that Higher Education Institutions need to embrace changes by aligning their institutional structure and operating methods to such needs. The key drivers for such demand are digitalization, demographic trends, internationalization, globalization, and learner-centricity. Thus, HEIs should rethink and reform their structures and operating methods that can help create diversified competencies which in turn solve the demand of the labour market and the society (Publications of the Finnish Government, 2021).

2.8 The Open Science Agenda in Higher Education Institutions

The OECD (2015) published a report under the theme of open science. Before going into details on the concepts of open science, it is necessary to define the terms related to opens science based on the report. As indicated by OECD (2015) report, formal definition of open science was still missing. The same report defined open science as:

"Efforts by researchers, governments, research funding agencies or the scientific community itself to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction as a means for accelerating research; these efforts are in the interest of enhancing transparency and collaboration and fostering innovation" (p.7).

The Open science ecosystem embraces different perspectives within itself such as open research data, open collaboration, open access, and open-source software (OECD, 2015). Open access refers to accessing scholarly works without restriction which can be enabled through institutions, databases, publishers, and scholars' websites etc. Open data is defined as data that can be accessed, used, and reused by anyone interested without restraints. In the same way open research data promotes usage of scientific works to anyone interested to access it (National Open Science Coordination, Federation of Finnish Learned Societies, 2021).

Within the same direction, in Finland, the national and executive plan of higher education and research community published by National Open Science Coordination, Federation of Finnish Learned Societies outline the policy components pertaining to open research data and methods. The objective of the open research data and methods policy is to enhance open science principles and facilitate opportunities for disseminating and reusing scholarly works. It outlines that "*research data and methods should be as open as possible and as closed as necessary*" (p.5). It also affirms that open access to research data and methods is affected by the EU regulation such as Open data directive which aims at providing legal framework to uphold openness of data. In line with this, the policy points out that the extent of openness of research data and methods should have strong justification.

Furthermore, the Finnish open policy identified potential risks associated with openness of research data within the light of good research data management practices. The issues regarding opening research data and methods can be seen from researchers, funding organizations, economic perspectives. From financial aspect, it needs considerable resources (i.e., financial, human, and technological) to manage the research data throughout its lifecycle. From the researchers' perspective, the degree of openness of data is one of the issues with respect to "*opening sensitive and confidential research data (e.g., personal information) and methods*" (p.8). Another concern for researchers is the issues related to legal protection such as IPR (Intellectual Property Rights), and contracts such as agreement between the funding organizations and the researchers. Therefore, the degree of openness needs to be aligned with the primary use of the data (National Open Science Coordination, Federation of Finnish Learned Societies, 2021).

From funding organizations and publishers' perspective, opening research data and methods is becoming a norm. Due to these requirements, applicants who are unable to open their data may fail to access the funds. The major inhibitors for opening research data are absence of technology or infrastructure, lack of necessary skills, legalizations and contractual restrictions, and research ethics (National Open Science Coordination, Federation of Finnish Learned Societies, 2021).

The policy developed by National Open Sience Coordination, Federation of Finnish Learned Societies (2021) also sets out objectives and action that Higher Education Institutions are required to execute. The objectives demand HEIs to develop data management plan for research and development (no later than 2023), apply operating models to rights, terms, and license pertaining to the use of research data (no later than 2022), document research and development project to support the FAIR principles (no later than 2022), and create storage solution, infrastructure, and services to promote good data management throughout the lifecycle of research data (no later than 2022).

The Finnish Declaration for Open Science and Research (2020–2025) stresses that in research and science, openness is essential factor that improve the effectiveness, quality, impact, innovation of research, scientific or artistic, in the research community. The main target of opens science and research agenda is to promote equality (Open Science Coordination in Finland, Federation of Finnish Learned Societies, 2020).

According to Kassen (2022), many HEIs were participating in different projects that promotes open data. The Finnish HEIs have been collaborating with both public and private sectors in developing numerous open data projects to advance data-driven innovations. Along with this, the HEIs have been generating extensive datasets pertaining to education, research, science, and technology. Afterwards, the datasets were shared with private, public, and non-governmental sectors to promote reusability of research data (Kassen, 2022).

2.9 Legislation in Finnish Higher Education Institutions

As public organization, the Higher Education Institutions in Finland comply to certain regulations and laws pertaining to data protection matters. As mentioned above, the Finnish HEIs are classified into two broad categories, namely, universities and university of applied sciences. Most common laws that Higher Education Institutions must follow are General Data Protection Regulation (GDPR), Data Protection Act, universities Act, and University of Applied Sciences Act.

As reported by DLA Piper (2023), although the GDPR entered in to enforce in 2016, it comes to effect on 25 May 2018 in all member states of European Union, including

Finland. In many ways, the regulation (GDPR) affects how HEIs handle, process, transfer (i.e., to non-EU countries) data. It also sets out rules for data processors and data controllers. The GDPR outlines multiple rights for data subjects in HEIs such as the right to access, rectify, erasure, restrict processing, data transfer the data to another controller, be informed regarding data breach, and withdraw consent.

According to the Finnish Data Protection Ombudsman Office (2023), the Data Protection Act (Tietosuojalaki) (1050/2018), supplements the EU GDPR which entered enforce in January 2019. It provides appointment and powers of supervisory authority regarding data protection issues. Among others, Data Protection Act sets out principles in relation to processing special categories of personal data (Section 6), processing personal data for academic, journalistic, artistic or literary expression (Section 27), legal basis for processing personal data for public interest (Section 4), and restrictions pertaining to data subject (Section 34).

Other regulations concerning the Higher education are Universities Act and University of Applied Sciences Act. Both acts provide guides regarding data protection such as how to handle sensitive material, how to use data for supervision and reporting, and how to access information related to studies or admissions. Beyond that the HEIs has their own regulations pertaining data protection to safeguard personal data.

2.10   The FAIR principle in data governance context

The abbreviation FAIR stands for Findability, Accessibility, Interoperability, and Reusability. Wilkinson et al. (2016) published article titled as "FAIR Guiding Principles for scientific data management and stewardship." The authors stated that the goal of FAIR principle is to promote good data management and increase re-usability of scholarly data by using machine capability with minimum human intervention. They also added that the FAIR data principles can serve as guideline to augment data reusability. The principles leverage the machine capacity (machine actionability) to find, access, and reuse research data. Wilkinson et al. (2016) argue that however humans have "intuitive sense of semantics" to interpret the contextual cues, they often face difficulties when there is lack of context in the data (metadata). In addition, humans lack to work on voluminous amount of data with necessary speed. These demanded that humans to rely on computational

power to collect, integrate, and maintain accuracy and increase reusability of data with several formats, and types (Wilkinson et al., 2016). Therefore, by overcoming those problems, different stakeholders can obtain benefit from the implementation of FAIR principle in research data. For instance, researchers can easily access, and reuse other researchers' data openly, and funding agencies can build long-term data stewardship program by utilizing research data. According to Jacobsen et al. (2020), the FAIR principles were explained as follows.

1. Findability: Digital resources should be easily findable by human ad machines.
2. Accessibility: In order to retrieve and access digital resources there should be explicit protocol and clearly defined mechanisms to authorize access.
3. Interoperability: It should be possible for machines to merge two or more related digital resources to provide richer information
4. Reusability: digital resources should be well defined to enable machines and humans in making decision pertaining to reuse, and credibility

2.11   Organizational roles and responsibilities in data governance

Previous studies indicate that data governance involves designating decision making right, roles and responsibilities, and accountabilities to obtain the maximum value from the existing data assets.

As mentioned by DAMA (2017), to implement effective data governance program, the roles and responsibilities of people involved in data governance program need to be well defined and understood. According to a report conducted by Plaid Consulting (2021) on data governance practices at 14 Canadian postsecondary institutions, the most common roles are data stewards, data trustees, data custodians, data guardians, and data users are presented as follows.

1. **Data stewards**: are individuals who are responsible for supporting data use by developing policies, plans, procedures at the university (Plaid Consulting, 2021). Hence, data stewards are responsible to ensure that institutional data assets are of high quality and fit for the purpose they are intended for (DAMA, 2017). According to DAMA (2017) data steward focus of work and their place may differ

from one organization to another. For instance, they can act as one of the following roles based the complexity of the organization: chief data stewards, executive data stewards, enterprise data stewards, business data, data owner, technical data stewards, and coordinating data stewards.

2. **Data trustees**: are individuals at highest rank of data governance who are accountable for what happens with and to university data such as institutional and research data.

3. **Data custodians**: are individuals responsible for safeguarding that policies in certain areas and processes are consistent with the institutional policies and procedures.

4. **Data Guardians**: are individuals responsible for data management activities such as data creation, storage, maintenance, cataloging, deletion, and dissemination at operational level.

5. **Data users**: individuals who access and use institutional data.

2.12 Data governance Frameworks

Many organizations develop developed data governance framework that fits to their organizational needs. As reported by PwC (2019), data governance framework involves developing a model to manage organizational data assets. Therefore, it enables organizations to establish guidelines and procedures pertaining to their data assets (PwC, 2019). The following section illustrate the most common data governance frameworks in academic research and industrial practices.

2.12.1 Data Management Body of Knowledge Data Management framework

The Data Management Association (DAMA) published the first edition of a guide to data Management Body of Knowledge in 2009. It has been used for various research about data management and data governance since then. The second edition of Data Management Body of Knowledge (DAMA-DMBOK2) was published in 2017 due to the significant development in data management field of study (DAMA, 2017). According to Okoro (2021), the data governance framework published by DMBOK has been used as a guide to conduct research related to data governance. DAMA-DMBOK2 explains that

due to the increasing adaptation of data governance in many organizations for standardizing structure pertaining to their voluminous amount of data and the need to consider data ethics along with a large amount of data collection that requires analytics, DAMA revised the existing frameworks. The framework is explained as follows.
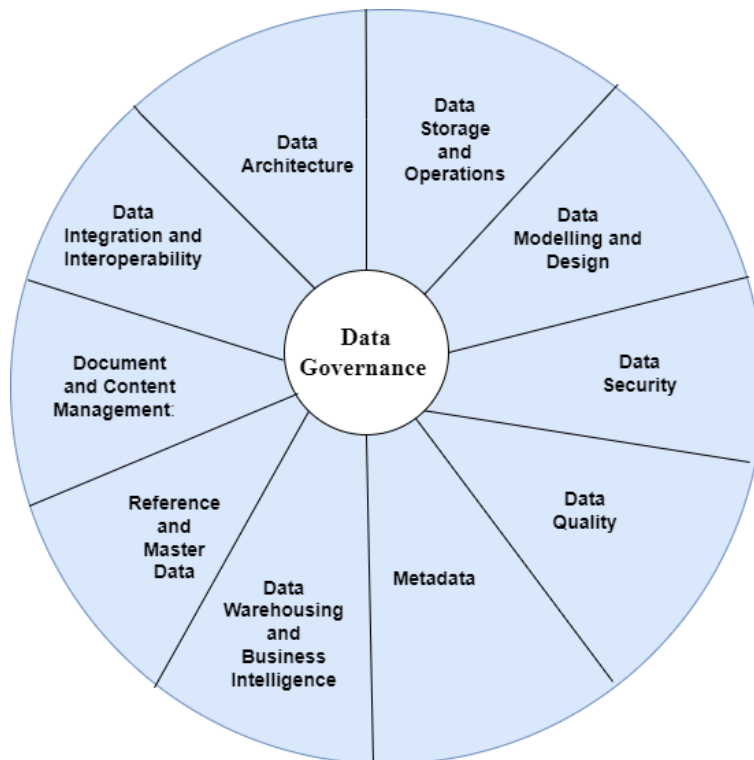
Figure 2 The DAMA Data Management Framework (DAMA, 2017)

According to DAMA (2017), data governance placed in the centre of data management activities to ensure the functions consistent and balanced. The framework consists of 11 (eleven) Data Management Knowledge Areas including data governance. These knowledge areas are described as follows.

- **Data Governance**: It concerns establishing standards, policies, and frameworks pertaining to decision rights over the data assets of enterprises.
- **Data Architecture:** It is related to the standardization and integration of organizational data assets which can serve as a blueprint for managing data assets by

making it a consistent organizational strategy. For example, designing enterprise data models.

- **Data Modelling and Design**: It is an essential component of data management that involves discovering, and documenting business data requirements in order to cut costs pertaining to data support and increase reusability.

- **Data Storage and Operations**: The main objectives of data storage and operations are to ensure the availability and integrity of data assets through the whole data lifecycle (from planning to disposal stages), and to manage data transaction performances. This knowledge area requires more technical skills.

- **Data Security**: It concerns maintaining and protecting the data asset of enterprises aligning with data privacy and confidentiality policies, regulations, contractual agreements, etc.

- **Data Integration and Interoperability**: It consists of processes regarding data consolidation and data movement among data centers, data stores, applications, and across organizations.

- **Document and Content Management**: includes planning, implementation, and control activities used to manage the lifecycle of data and information found in a range of unstructured media, especially documents needed to support legal and regulatory compliance requirements.

- **Reference and Master Data:** involves maintaining and reconciling data assets to avoid inconsistencies in data structure and data values that could possibly be caused by multiple mergers and acquisitions.

- **Data Warehousing and Business Intelligence:** it consists of planning, implementing, controlling organizational processes in order to oversee organizational data and to ensure the organization obtain value from their data assets through analysis and reporting.

- **Metadata:** it consists of planning, executing, and controlling activities to ensure that organizations can access to integrated and high-quality metadata**.**

- **Data Quality:** it involves of planning and implementing data quality management techniques and metrics to improve, evaluate, and measure whether the data fits for specific purposes or uses.

2.12.2 The IBM Data Governance Maturity Model

The IBM Data Governance Maturity Model was developed in 2007 to ensure data quality, availability, and integrity so that stakeholders can understand when and how to start data governance solutions to the existing problems (IBM,2007). According to (Sunil Soares, 2010), the data governance maturity model of IBM can also assist stakeholders to make an effective strategy. The framework consists of 11 categories namely data risk management and compliance, value creation, organizational structures, and stewardship, policy, data quality management, information lifecycle management, information security and privacy, data architecture, classification and metadata, and audit information logging and reporting. The 11 categories of data governance are further divided into four interrelated groups namely outcome, enablers, core discipline, and supporting disciplines. The below figure illustrates the IBM data governance maturity model.
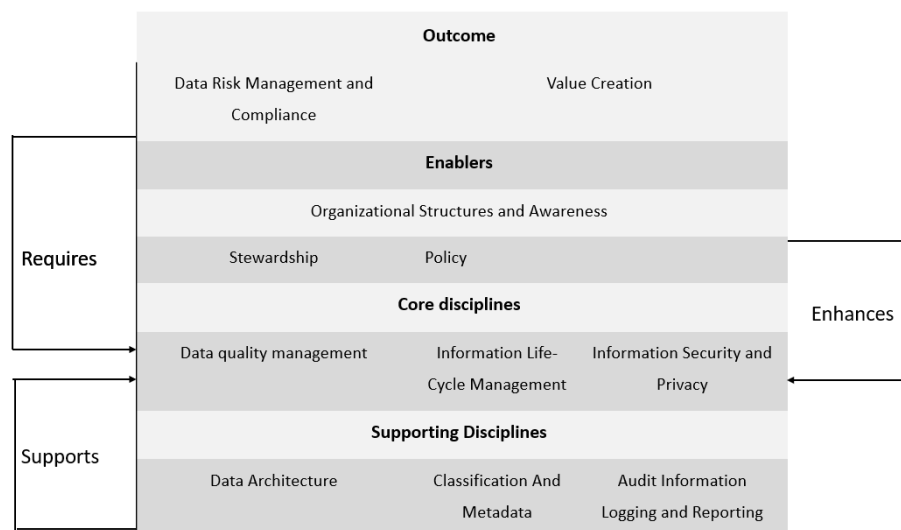
Figure 3 The IBM Data Governance Maturity Model (Sunil Soares, 2010)

2.12.3 The Data Governance Institute (DGI) Data Governance Framework

The Data Governance Institute (DGI) data governance framework is aimed at ensuring clarity, establishing accountability, reducing costs, defining the scope, maintaining focus, and defining measurable successes (Thomas, 2020). Thus, The DGI has developed a data governance framework that consists of 10 components within three major focus areas. The components are Mission and Vision, Goals, Governance Metrics, Success Measures,

Funding Strategies, Data Rules and Definitions, Decision Rights, Accountabilities, Controls, Data Stakeholders, Data Governance Office, Data Stewards, Proactive, Reactive, and Ongoing Data Governance Processes. Those components are categorized under three focus areas: people and organizational bodies, rules and rules of engagement, and processes (Thomas, 2020).

2.12.4 The PwC Enterprise Data Governance Framework

According to PwC (2019), even though organizations heavily invest their resources on their data value chains , they often face difficulties in deriving tangible value or benefit from data asset due to the lack of data governance. Therefore, using standardized data governance framework can significantly decrease data management costs, boost their ability to manage risks, and ensures compliance with regulatory. The objective of PwC's holistic data governance is to maintain the existing and potential data ecosystem organizations and to help them to adapt data driven programme. The PWC's Data Governance Framework consists of five components. The definition of each component is presented as follows (PwC, 2019):

1. **Data governance strategy**: It is the initial point for organizations whereby they define their business goals and plan to establish metrics to track their progress and achievements pertaining to the implementation of data governance.

2. **Enterprise data management:** In this component, essential areas pertaining to organizational data ecosystems such as data integration, metadata, data models and architecture, business analytics, and reference data management defined and covered.

3. **Data lifecycle management:** It is about managing the data flow throughout its lifecycle, from the data capturing until discarding. So, organizations can save cost, protect data assets, increase efficiency, and boosts data integrity.

4. **Data stewardship and core functions:** This component covers the decision on prioritizing core areas and setting Key Performance Indicators (KPIs). It also consists of operations such as data lineage, data privacy, data security, data access and control, metadata management, data quality, and data quality.

5. **Data governance enablers:** In this component, the enabler of data governance namely people and culture, process and operating models, and tools and technology help to assess the existing processes and maturity level of their data governance, identify gaps pertaining to data quality, design pertinent solutions and roadmap to implement data governance.

2.12.5 KPMG data governance Frameworks

According to KPMG International (2018), data governance helps organizations to define how they should manage their data assets and how that decision-making should be operationalized. As reported by KMPG, "*this calls for an appropriate authority model to manage data functions*" (p.3). KPMG International (2018) proposed a data governance framework consists of four key components, namely data stewardship, data ownership, data policies, and data standards. Figure 4 illustrates the elements of the framework.
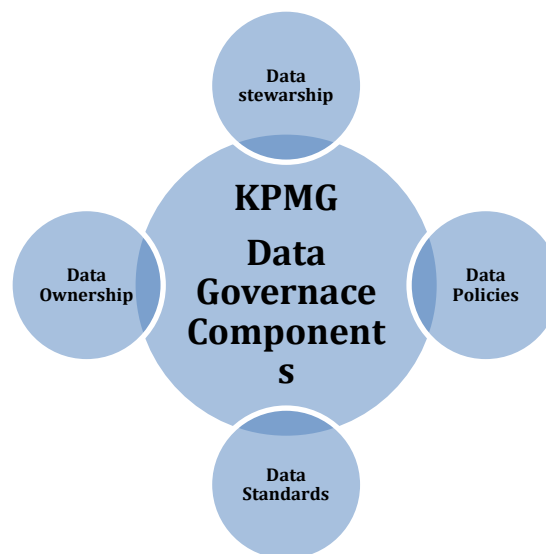
Figure 4 KPMG Data Governance Framework (KPMG International, 2018)

Definitions of each data governance component shown in Figure 4 are described as follows.

1. **Data Stewardship:** It encompasses setting accountability to manage institutional data assets. Data stewards play vital role by ensuring enterprise-wide data quality, security, and accuracy.

2. **Data Ownership:** It concerns with placing responsibility for creation of data, enforcing business rules, and defining data use throughout the organization.

3. **Data Policies:** Data policies are the rules that organizations use to enforce data access rights, compliance with regulatory frameworks and manage their data assets.

4. **Data Standards**: Data standards are set of specification, rules and criteria that organizations utilize for governing their organizational data assets and defining data lifecycle.

2.12.6 Khatri and Brown's The Five Decision Domains Data Governance Framework

Unlike previously mentioned data governance frameworks, Khatri and Brown (2010) proposed a framework that focuses on decision-making rights and locus of accountability regarding organizational data assets. The authors claim that to design an effective data governance framework it is essential to *"step back from day-to-day decision-making and focus on identifying the fundamental decisions that need to be made and who should be making them"* (Khatri and Brown, 2010, p.148). Therefore, data governance-related decisions are more strategic than operational. Khatri and Brown's (2010) Data Governance Framework consists of five (5) decision domains which are derived from Weill & Ross's (2004) IT governance decision domains. The IT governance framework that Weill & Ross's (2004) proposed also includes five decision domains such as IT Infrastructure, IT principles, IT architecture, Business application needs, and IT investment and prioritization. In comparison, Khatri and Brown's (2010) data governance framework comprises five interrelated decision domains, namely data principles, data quality, metadata, data access, and data lifecycle. The following figure the data governance decision domains.

Data Principles		
Data Quality	Metadata	Data Lifecycle
	Data Access	

Figure 5 Data governance framework (Khatri and Brown, 2010)

2.12.7 Summary of the data governance frameworks

The data governance frameworks mentioned above were widely applied in industries academic research. There are similarities and differences between the frameworks based on the scope they cover, level of complexity, level of maturity (i.e., due existing data governance practice), and type of organizations. For instance, the PwC Enterprise Data Governance Framework and the IBM Data Governance Maturity Model BM is more suitable for organization who already implemented data governance program. Whereas the DGI data governance framework fits for organization who plans to develop data governance program form the beginning.

In other hand, data governance model such as DAMA-DMBOK-2 combines data management and data governance into one framework. It covers wide-raging areas from both data management and data governance components which more suitable for large organizations. Essential components such as metadata and data quality management are missing in the KPMG data governance framework. In contrast, data framework developed by Khatri and Brown (2010) provides comprehensive components(i.e., decision domains) to assess decision-making rights and locus of accountability concerning organizational data asset.

2.13 Conceptual framework of the study

For this study, the data governance framework developed by Khatri and Brown (2010) was selected because the framework is well recognized in academic research and provide comprehensive and simplistic decision domains to study data governance practices in HEIs. In addition, as Khatri and Brown (2010) state the framework can assist researcher to emphasis on significant data governance issues. These most important issues are covered under the data governance decision domains. The other reason is that the framework associates IT assets and data assets as that framework itself is derived from IT governance. Because IT infrastructures play significant role in production, storage, analysis modelling, archiving , and organizing data assets. Thus, it links IT governance with data governance decisions. Previous studies also indicated that the data governance issues in higher education are understudied. Therefore, this framework can assist the researcher to explore the challenges and opportunities of data governance practices in

HEIs. Khatri and Brown's (2010). The five Data Governance decision domains are explained as follows.

1. Data principles

Data principles is about clearly identifying and defining the role of data as organizational asset. This domain is gives guidance for the rest of data governance by delineating conditions for business use of data. For example, by setting a boundary pertaining to business use of data , data quality standards, metadata data interpretation procedures or methods, data accesses requirements , procedures and policies can be formulated. It also impacts the decision the data lifecycle such as data creation, retention, and deletions. By revealing the role of data as an organizational-wide data asset, data principles domain focuses on the following decisions pertaining data governance.

- Formulating suitable policies, procedures, programs, standards, and guidelines
- Establishing foster opportunities for sharing and reusing data assets
- Distinguishing the required behaviours both for IS professionals(e.g., data stewards) and business users.
- Identifying the influence of regulatory environment that could affect business use of data.

2. Data quality

According to DAMA(2017), to gain the required value from the data asset the data itself should be reliable and trustworthy (i.e., should possess high quality). The aim of the data quality is to determine the requirements of data use purposes or intentions. Consequently, the context of intended use of data should be defined. Based on data quality measures or attributes such a timeliness, completeness, accuracy, and credibility it can be synthesized with the end goal of data use. In addition, data quality decision domain also specifies issues related to means (i.e., programs) for communicating data quality, and mechanisms to monitor and evaluate data quality etc. By answering to these probes, data quality decision domain provides a direction or guideline for the interpretation of data, i.e., for metadata decision domain. Thus, data quality plays " pivotal role in the effective governance of data assets "

3. Metadata

Before going into metadata decision domain, it is important to define the term since term is confusing due to vendors overuse and bias (Ladley, 2012). According to Hoffer et al. (2016) meta data can be defined as data that portrays the properties and context of end-user data. It outlines, "*what the data is about*" and gives coherent and concise description of data which in turn ease the task of interpreting the meaning of the data itself (Hoffer et al., 2016, (p. 42-43); Khatri and Brown, 2010). There are four major types of metadata such as user meta (it includes annotations such as user history user and user preferences), physical metadata (it consists of information regarding physical storage of data) , domain-specific (it consists of descriptions data regarding specific unit), and domain-independent (it involves information regarding data authorization and data creator(s)).

The core focus of metadata decision domain is to ensure the decide up on the semantics of data to make it easy for end-user to interpret it. Hence, it is necessary to set metadata standards for effective use of the data asset. Besides, metadata domain also needs to allow to manage changes as the requirement business use changes in the future .

4. Data access

Data access decision domain involves determining data access requirements. Thus, it covers data access standards, policies, program and procedures, data auditing, dissemination of data security and education, risk assessment aligned with regulatory compliance, and backup and recovery program. As stated by DAMA (2017), data auditing helps to document data lineage which in turn can be used for validation and historical tracking.

5. Data lifecycle

This domain give emphasis to how data is produced, how long it should be kept and discarded. As mentioned by Khatri and Brown (2010), it helps organizations to develop a mechanism *"to map usage patterns to the optimal storage media, thereby minimizing the total cost of storing data over its life cycle"* (p.151). In this domain's decision pertaining to data inventory, legal and regulatory compliance that affect the data archiving and retention , and description of data lifecycle, from production to deletion need to be considered. For instance, developing data taxonomy can enhance the management of data lifecycle.

Data Principles		
• Data ownership and data stewardship • Business use of data • Data as an institutional asset • Communication mechanisms • Opportunities for sharing and reusing • Influence of regulations and laws on data use • Data value/use challenges		
Data quality • Data quality attributes(accuracy, Timeliness, completeness, and credibility) • Data quality tools and plan • Monitoring, and cleaning data quality • Data quality audits • Data quality Challenges	**Metadata** • Program for documenting semantics of data • Metadata definition and modelling • Plan regarding metadata up to date(types) • Data dictionary	**Data Lifecycle** • Data inventory and archiving • Guidelines for defining data lifecycle (production , retention, discarding) • Influence Legislation on data archiving and retention
Data Access • Polices and procedure that monitor data access • Internal procedural control • Risk assessment plan • Consistency with GDPR • Data security education and awareness mechanisms • Backup and recovery plan		

Figure 6 The modified Data governance framework (Khatri and Brown, 2010)

3 METHODS

3.1 Research Purpose

According to Saunders et al. (2016), most literature on research methods classify the purpose of research into three categories namely, descriptive, explanatory and exploratory. However, the same authors indicate that depending on the topic, objective of the study, and research questions, a research project may have more than one purpose.

Exploratory research is a useful tool for posing open questions to learn and build understanding about a particular subject of interest. The exploratory research questions will often start with "What" or "How." It is also possible that the questions that researcher pose during data collecting to investigate a problem or phenomena begins with "What" or "How" questions. Exploratory research is advantageous when a researcher is uncertain of the specific nature of a problem. The methods for conducting exploratory research vary. A literature search is one of them, along with interviewing "experts" in the field, conducting focus groups or in-depth one-on-one interviews (Saunders et al., 2016).

This research is exploratory research in which semi-structured interview was conducted with experts from different areas to gain insights in relation to challenges and opportunities of data governance practices at HEIs. It was chosen because very limited studies have been done pertaining to data governance since the research area is relatively new.

3.2 Data collection methods

Saunders et al. (2016) defines interview as *" a purposeful conversation between two or more people, requiring the interviewer to establish rapport and ask concise and unambiguous questions, to which the interviewee is willing to respond, and to listen attentively. Essentially it is about asking purposeful questions and carefully listening to the answers to be able to explore these further" (*p.388). Interviews are used to collect data from primary sources. An interview, according to Krishnaswamy and Satyaprasad (2010), is a systematic conversation between an interviewer and an informant that aims to gain information relevant to the specific study. They add that the interview method is extremely useful for gathering in-depth and detailed contextual information from

informants, which is accompanied by reasons. In other words, interviews can assist researchers in obtaining valid and reliable data pertinent to the research questions and objectives (Saunders et al., 2016).

Interview can be classified into three categories namely structured interviews, semi-structured interviews, and unstructured or in-depth interviews. The selection among these categories should always depend on the research questions and objectives. In structured interviews, interviewers provide the same questions or identical set of questions to all the participants using the same intonation and order to avoid bias (Krishnaswamy & Satyaprasad., 2010; Saunders et al., 2016). They are also called 'quantitative research interviews' since they are used to gather quantifiable data (Saunders et al., 2016). In contrast, there is no standard set of questions in semi-structured interviews; rather, the researcher develops key questions to be covered based on the specific themes (Saunders et al., 2016). Researchers could take notes or record the audio to gather the data. Another type of interview is unstructured interview where respondents are given minimum guidance to discuss freely on a particular topic (Krishnaswamy & Satyaprasad., 2010).

For this study, semi-structured interview was selected because the researcher formulated research questions based on predefined themes. Furthermore, since this research is qualitative research, it is suitable to utilize semi-structured interview to gather the necessary data. The interview questions for this research were formulated based on Khatri and Brown's (2010) the five decision domains of data governance. The interview questions were developed around five major themes: data principle, data access, data lifecycle, data quality, and metadata. To formulate the research questions numerous literature, guidelines, laws, regulations, and acts have been considered. After the interview questions were formulated, the researcher sent invitation email to experts who are interested in the topic of the study and have been relying on data for decision-making.

3.3 Data sources

As indicated by Krishnaswamy and Satyaprasad (2010), data sources can be divided into two categories, primary and secondary. In primary data sources, researchers gather data that have never been collected before, while secondary data sources are composed of information that has been gathered and compiled for another reason (Krishnaswamy and

Satyaprasad, 2010). Primary sources of data can be collected through different methods such as interview, questionnaire, and mailing. In contrast, secondary sources of data can be collected from reports, statistical statements, etc.

In this study, data from both primary and secondary sources were collected to attain objective of the study and answer the research questions. Experts from different fields were interviewed through semi-structured interview questions. As secondary data sources, documents related to international, local, and European Union (EU) laws and regulations pertaining to personal data processing were considered. In addition, to understand the existing data-related practices guidelines that reflect the data and information management plans at HEIs were introduced as secondary data sources.

3.4 Sampling Techniques

Using the right sampling technique is crucial for achieving research objectives and addressing the research problem. It will frequently be difficult for a researcher to obtain all the potentially helpful data that is available for the research questions and objectives due to time, money, and access constraints (Saunders et al., 2016). Hence, by choosing proper sampling technique, a researcher can limit the amount of data needed to be gathered by focusing on data from a specific subgroup rather than all potential cases or population elements. Moreover, data collection methods selected for specific research also affects the sampling techniques.

According to Saunders et al. (2016), sampling techniques can be classified into two major categories namely, probability and non-probability sampling. The former is mostly applied in survey research strategies where it is necessary to make inferences about the population from the sample to satisfy the objectives and find answers to research questions while the latter is used when the representative sample of the population may not be appropriate to answer the research problem. The underlying assumption regarding probability sampling is that samples from the population are selected from the sampling frame while in non-probability sampling does not consider the theory of probability (Krishnaswamy and Satyaprasad., 2010; Saunders et al., 2016).

Due to the nature of the research topic including research questions and chosen data gathering techniques, the non-probability sampling strategy was employed for this study. The goal of the study is to explore the difficulties and possibilities associated with data governance practices in Finnish higher education institutions. The researcher chose a non-probability strategy to comprehend the data governance practices in higher education because there are very few resources available in terms of literature. Also, subject matter experts' insights were required to attain the study objective.

To include all the aspects of data governance decision domain as outlined by Khatri and Brown (2010), Finnish HEIs who collaborate and share resources with each other were selected purposively. Therefore, the researcher chose heterogeneous purposive sampling to choose sample components based on the previously indicated specified criteria. Experts from different areas such as administration, IT, and library at HEIs were selected using heterogeneous purposive sampling technique. Since the research framework consists of specific themes (Saunders et al., 2016) that should be addressed by experts from different areas to provide answers to the research problems, for this study, heterogeneous purposive sampling technique was suitable.

As a result, professionals from different units at HEIs were considered (i.e., based on their exposure in using and managing data to make decision and establish strategies). Out of fifteen interview invitations, only eight professionals from two HEIs (i.e., from both University and University of Applied Science) were interested to have interview. The researcher did not receive response from the third HEI experts. One of the eight interviews was utilized for pilot testing while others were retained for further analysis. Therefore, the total number of participants in this study was seven. The researcher conducted the interviews both online (using zoom meeting), and face-to-face depending on the preference of the participants.

3.5 Data presentation and analysis

To transcribe and edit interviews, and writing the results of this study, MS-word (Version 2302) was used. According to Alhojailan and Ibrahim (2012), thematic analysis is one of qualitative analysis method which often used to analyse categories and present themes related to the gathered data. It could be used when the aim of study is to comprehend the existing practices of particular situation. It also allows the researcher to grasp the

relationship between the collected data and concepts or themes (Alhojailan and Ibrahim, 2012). Since the chosen framework in this study consists of specific themes, thematic analysis is used to examine the themes in this study. The research findings were displayed using texts, and tables.

3.6 Data preparation and analysis tools

Data analysis software are beneficial to analyze, label, and organize qualitative data. In addition, data analysis software also aids to categorize qualitative data into similar themes (Alhojailan and Ibrahim, 2012). Based on the conceptual model chosen for this study, the themes are categorized using NVIVO (version 1.5.1). As mentioned by Linneberg and Korsgaard (2019), there are two main coding approaches in qualitative exploratory research, namely, inductive, and deductive coding. The former is useful when a researcher is unable to access underlying theoretical concepts relevant to his/her study whereas the later emphasized on coding on themes that are drawn from the existing literature (Linneberg and Korsgaard, 2019). In this study, deductive coding was employed because the themes for the study were drawn from Khatri and Brown's (2010) the five decision domains of data governance.

4 ANALYSIS AND INTERPRETATION OF RESULTS

This section describes the analysis of the semi-structured interviews is described where the first subsection of this chapter outlines participants information. In subsequent subsections, the five decision domains of data governance, namely data principles, data quality, metadata, data access, and data lifecycle are discussed accordingly.

4.1 Participants

To present the results of this study, the participants in the interviews were categorized into three broad categories based on their roles, occupation areas, and responsibilities. These categories are librarians, administration, and ICT. The HEIs libraries share the same resources, both printed and electronic. The research expert also works with the library for data management, and research data management under the open science theme. Table 1 summarizes the demographic data and duration of the interviews for each participant.

Table 1 Participants information

No.	Job title	Occupation area	Duration of the interview(in minutes	Date of the interview(In months)
1	Person A	Administration	58	March, 2023
2	Person B	Administration	64	March 2023
3	Person C	Library (research services)	66	March 2023
4	Person D	Library	82	March 2023
5	Person E	Library	56	March 2023
6	Person F	Administration	52	March 2023
7	Person G	IT Support services	67	March 2023

4.2 Data Principles

According to Khatri and Brown (2010), data principles involve defining data as an organizational asset and identifying business use of data. This section discusses the elements of data principles decision domain, such as data ownership and stewardship, business use of data, data institutional asset, influence of regulatory environment on business use of data, opportunities for data sharing and reusing, and communication mechanisms pertaining to business use of data are discussed.

4.2.1 Data stewards and Data Owners

Respondents from the administration described that they have data stewards in different units. From the research group, Person B (Administration) expressed that study advisors act as data stewards.

> *"Yes, of course. There are study advisors in that sense, I am thinking of the data as course information. That's data. I see it because they are responsible that teachers hand in like all the information about their courses, learning goals, and when and where the course should be taught. So that is the data stewards' responsibility to ensure that data is available in the study data system. And it then is viewable and readable for students in the study guide".* **Person B, March 2023**

Person F (administrative staff) from Continuous Education stated that planning officers are responsible for guiding students and other customers (i.e., non-degree students from the community interested to take courses from HEIs). According to this participant, the planning officer ensures that all the stakeholder needs are fulfilled in their unit. However, their unit has no dedicated data steward.

On the other hand, librarians (Person D and Person E) stated that the role of data steward is defined under research data management which incorporates a working group under the theme of open science and research. There is a data steward position or role in the research data management team. This is evidenced by the following statement from a research services expert.

> *"I am the data steward. The one who has the data steward function at the university. And the closest you would get to that function. But there are also other specialists that sit and join me in the team because we have a research data email. I am responsible for answering all the questions. But some questions are more concerning GDPR. And then I contact the GDPR lawyer".* **Person C (Library-Research services), March 2023**

Person D also indicated that the project on libraries requires data stewards due to the increasing amount of data they store and the sensitivity of data. Person G (ICT) said that the data steward role is well-defined in open data support(i.e., research data management). The respondent also confirmed that the information security officer at HEI had been acting as a data steward before the open data project.

Person G (ICT) explained that they have similar roles to data stewards in other departments such as human resources and finance since they have been managing unit-specific data assets as evidenced by the following quote:

"Perhaps, one could say that functions such as human resource and finance are sort of data stewards. I mean, that's because they handle, and they tell the rest of the organization how to handle specific datasets. So, these are perhaps also done, and this has also, of course, been in existence for a long time". **Person G, March 2023**

Among all participants, the question related to data ownership was only addressed by Person A. Person A from the administration described that the administration has different sections, and each section has its data owners. The same respondent said that:

"HR specialists have a lot of data about personnel, teachers, researchers, and those who are looking at the student offices. Yes, they of course own data. however, we don't have data specialists, but we all who are doing this job know and must know how we are handling all the data. And we have also a lot of policy documents.We all have been working together to make them work". **Person A, March 2023**

According to DAMA (2017), subject matter experts can be data stewards even though their position is not clearly labelled through a data governance program. The result of the study revealed experts from different units play the role of data stewards. However, data steward role was clearly defined in research data management (research services) unit which operate within the national open science program.

4.2.2 Data as an institutional asset and business use of data

Concerning business use of data, all participants responded that data is crucial to develop reports, assessing projects and follow-up research data management progresses. Person A stated that *"of course, it is important that we have the data that we need for our job "*. According to Person B (administration), data are crucial assets to provide services to customer groups, develop statistics for the management as evidenced by the following quote.

"It is crucial. You know, knowledge is everything and without relevant data, it is hard to do to provide services for our customer groups. For instance, a lot for

academics or especially the management ask for statistics quite often". **Person B, March 2023**

As evidenced by Person D, data is essential to support research data management projects and conduct research.

"I would say it is kind of important because we have very many projects going on. And it's also said that students should learn in some phase of their student studies how to make a data management plan. I don't think it is fully implemented yet, but the national policy the Finland says that students also should know about it. Because many of our students do small data collection such as interviews or questionnaires, these kinds of things. So, yes, it's a good thing if they know something about it". **Person D, March 2023**

Person E showed a different perspective on data assets. The participant indicated that data have economic value and if it is managed well, it can speed up organization processes in a better way.

"Data is becoming increasingly important for everything that we do in our organization. And we are right now developing a system to collect organizational data so that we can store collections because we need to use it to take care of our responsibilities in a smarter way. It is an asset that worth money because we can use it to do things smarter. So, it is very important". **Person E, March 2023**

According to administration professionals, personnel data, students' data, study data, finance data, and research-related data are essential to generate reports and statistics. Thus, it can be determined that administrative experts use internal, external, and research data to make strategic decisions as evidenced by the following quote :

"We need a lot of data to generate statistics and reports. For example, we generate reports regarding which subject or programme is not doing so well. Why specific faculties don't have a lot of students. If they have a lot of students, we also make report why they have a lot of students. Why are the students not doing their master's degree or thesis? Based on the data we obtain and analyse; we make strategic decisions to support faculties. Also, we want data about researchers. For instance, we ask question such as which of the programme has a lot of researchers, but they don't publish articles? Because it is important to get funds from donors . Things like this". **Person A, March 2023**

Person G (ICT) defined data as *"extremely important"* asset because HEIs *"live and breathe data"* to perform different activities related to education, research, operational support, finance, and human resource. In the ICT department data had been used to handle security issues such as data breaching, phishing, or misconduct or misuse of institutional systems.

The participants were asked whether they consider data as an organizational asset. All participants responded that data is crucial asset for their organization. Person A (administrative staff) explained that data can be used to evaluate the correlation between numbers of students and teachers to generate reports pertaining to the student-teacher ratio and research output ratio (per faculty or per academic staff). Therefore, they would be able to decide on whether the faculty should continue or close by considering the minimum number of teachers in each program. Another importance of data asset for administrative experts is that it can be used to apply for accreditation from renowned institutions.

Person B agreed that data is an organizational asset that helps to make a strategic decision. To do so, *"you have to have current data, of course. I mean, you cannot really plan ahead and make qualified decisions unless you have the right basis"*. This indicates that relevant data is an essential asset that can be used for planning and making strategic decisions. Furthermore, Person B specified that *"we need a lot of data. So, we do our planning from the data that we had"*.

Additionally, Person F (administration) affirmed that data is a powerful asset to identify the needs of customers (i.e., degree students, non-degree students, and HEI staff) if it is utilized and analysed wisely using advanced analytical tools as evidenced by the following quotation:

> *"Data is kind of everywhere. And, and I think we are using all different types of data daily, even though we are not maybe even thinking about that. So, it's present all over and still. Even though that's most maybe not your question. So, I think you also have lots of type of data that we could use more wisely and analytically, even though we are not actually doing it now. ...So, for example, if you think about the data that we get from the participation of these open University courses, for example. So, I'm not quite sure that we could use more advanced tools for that, for learning analytics for example to facilitate learning and maybe analyze the need*

for different courses and so on. So, we are doing that. But, but I'm, I'm sure that there is potential to do more". **Person F, March 2023**

Furthermore, Person F (administration) stated that every activity that their unit involved in (i.e., continuous education) are connected to *"some data sources".* Besides, all administrative experts acknowledged that they have been utilizing their data assets to evaluate and estimate their performance regarding their current and future state.

All respondents from HEIs library agreed that data is valuable asset that benefits their institutions to deliver services appropriately and develop required reports as indicated by the following quotation:

"Yes, it is an asset. In more ways than I just said. Well, we are service organization. The university is our customer, in a sense, the one who orders services. And we also need to use the data to show that we do what we promised. So not only be smart within our organization, but also for more reporting purposes". **Person E, March 2023**

Moreover, Person E also affirmed that the library stores valuable cultural archives such as old photographic and manuscript materials. These cultural archives are currently in analog format but there is undergoing project aimed at transforming them into a digital format which in turn changes the resources into data. Therefore, cultural data could be utilized as a secondary data source for (further) research purposes as indicated below:

"But also, we have cultural archives which are very valuable and such as photographic material, but also manuscript material. And these are now in analog format, but when they are digitized. And this is something that we are working on a lot. They also become data. Here's another area of development for us now. We need to think about how to use and how we create and produce them. So, they become cultural data. For researchers, it becomes research data". **Person E, March 2023**

An interesting argument was raised by Person E (library) that if the data were co-created or owned by funders or external bodies, it will be difficult to call it an asset for the organization. However, the same respondent stated that *"But, they are valuable because they help us do what we should do".*

In general, all participants agreed that data is an institutional asset which plays significant role in making strategic decisions and developing plan, generating reports, conducting research, understanding customers, tracking misconduct, measuring performances, and providing services.

4.2.3 Communication Mechanism

According to the respondents from the administration, reports are a common mechanism for communicating business use of data on an ongoing basis. Specifically, reports regarding admission, gradation, and enrolment contain subjects such as how many students graduate each year, how many students are enrolled, how many credits have been taken by students in each faculty, why the number of students is dropped with a given faculty, how many students choose specific faculty. In addition, Person B from the administration mentioned that data could also be utilized to render comparisons among academic faculties performances.

Person F (administrative staff) also asserted that annual reports which consist of Key Performance Indicators (KPIs) and metrics were their standard tool to communicate business use of data. Also, their unit was attempting to benefit from existing data to develop quality improvement processes such as the PDSA Cycle (Plan-Do-Study-Act). For librarians, the business use of data was addressed through email. However, since most of the data that the librarians are concerned with is public and they gave less emphasis on communication issues. Typically, librarians communicate with user via the library rules.

Person E from the library responded that annual reports, statistics, and dashboards were the major tools they used to communicate business use of data as transpired from the following quotation:

> *"Well, we have, for instance, when we do our annual reports. It's a way of following to see what things are happening there in certain fields of expertise… it's a way of seeing the changes. And I think that there's much less of certain types of loans. So, there's much more of certain types of trends…. Well, the mechanisms, we are working on creating internal dashboard. But before we do that, we have to have sensible, reasonable, good data that actually helps the organization in its tasks. So, we first have to think about what we need and then we can put it out there. But the mechanism… certain numbers are asked for by the university board, and we submit*

that. And then our internal reporting is also such that some of sort of communication mechanism". **Person E, March 2023**

As Person G (ICT) stated, the most common mechanisms for communicating business use of data in the ICT department have been performed through policies, guidelines (instructions), courses, workshops, lectures, and email notification regarding data protection and security.

Generally, most common mechanisms for communicating business use of data a HEIs are guidelines and policies, emails, courses, workshops, lectures, reports, statistics, and dashboards.

4.2.4 Data sources

Alongside understanding the business use of data, identifying sources of data augments HEIs monitoring the data flows. According to Jim and Chang (2018), HEIs obtain data from internal (in-house), external (third party applications), and research or publications.

All administrative staff stated that their main sources of data were both internal and external but also publications. Person B (administration) described that "o*ur data asset is seldom research data, it's usually internal data. But we sometimes need external data like, like national data to know what is going on within the same field or within higher education institutions".* Alongside internal and external data sources, they also indicated that research data have been utilized to examine the research output in different faculties (i.e., number of publications per teacher or researcher) using data from the existing research information system. In addition, during COVID-19, students' data were used to examine the impact of COVID-19 on students' performances.

In contrast, librarians, including the research services, revealed that their sources of data stemmed from internal, external, and research. As evidenced by Person D , they collect and use data stemmed from publications (i.e., metadata), library data (loans and textbooks), cultural archives (old photographic and manuscript materials), personnel data, finance data, and project data:

> *"The publications are given to me by the researchers themselves. We actually collect them through JUSTUS[3], a system for Universities of Applied Sciences. There you can find a lot of data about publications. They are then open and free to look at from Research.fi. So, there you can see all publications collected in Finland by organization and year. So, these data are open for it does not contain anything that's sensitive because it's actually material that you can get from other places as well. There are many articles over there. You don't have to worry about sensitivity. But we have students from social and health care who might collect data that's sensitive, but I think they are aware of it. What to do. For example, they can't make films about young children".* **Person D, March 2023**

Person E (Library) also confirmed that the data is created not only internally, or externally but also from research as well. The same respondent elaborated that data could be co-created through collaboration with external parties. For instance, from the aspect of research data, data is co-created through data collection techniques such as an interview. It could also be generated by the apparatus from the laboratory of HEIs internally. Then ownership of the data may be affected if it involves external parties such as funders and collaborators because it raises issue of Intellectual property rights. On the other hand, organizational data (i.e., internal library data) are produced from HEIs activities and the contents of data may vary. For instance, it may be about personnel, book loans, number of collections and types of materials. Those data are used to generate reports and statistics which are later used by the national statistics of university libraries and HEIs management or boards.

Person G (ICT expert) said that data could stem from internal processes, external sources, or from research works. However, the expert specified that their unit was only concerned with the ownership and content of the data itself (i.e., whether the data is personal or whether it is guarded by regulations) as evidenced by the following quotation:

> *"I don't see it as much as from which process necessarily the data is stemming. I want to see it as which type of data is it? I mean, what information does this data contain? Is it personal information or is it a sort of guarded by laws? What is this agreement with the research partners or someone else which governs sort of how to protect data? So, I see it like this. But I mean, when handling the data and when processing the data, it all boils down to what data? What information does the data*

---

[3] JUSTUS provides publication information storage services. JUSTUS - Publication Notes Storage Service - CSC Company Site

contain. Also, I want to understand who owns the data, whose data is it, and who is responsible for the data". **Person G, March 2023**

Person F (Administration-Continuous Education) stated that their data sources originate from internal (human resource, students data system and finance), external (from the community, and job market), and projects (research). In this participant's view, external data, concerning job market forecasting and trends were used to examine competence needs. As a result, the unit, continuous education, can develop courses and training considering highly demanded skills in the market and internal data from their institution were used for financial planning, assessing staff and students' skill gap, and operational efficiency.

All administrative expertise indicated that their data source mostly internal and internal, whereas all respondents at HEIs libraries stated that their data sources can be internal, external, and research data. Other interesting aspect was raised by ICT saying that the concern of their unit is the content and ownership of the data itself rather than the sources. As a result, the ICT team could assess whether the data is sensitive or protected by the laws.

4.2.5 Opportunities for Data Sharing and Re-using

Participants from the administration stated that they share data both internally and externally. Only a limited version of data that is necessary for collaboration is shared with other HEIs or external parties. Similarly, Person F believes that most of the projects running within the continuous education unit needed data sharing with government bodies and funding organizations. According to the response of administrative experts, laws and regulations regarding data processing and protection have a significant impact on data-sharing activities. They mentioned that GDPR has an impact on data sharing and reusing. All administrative stated that they adhere to all the laws and policies and take corrective measures if something goes wrong as illustrated by the following quote:

"We really think that we are following all the laws and the policy document that we have. We think it is important. And sometimes if we see that something is not as it should be then we try to fixit". **Person B, March 2023**

All respondents from the HEIs libraries acknowledged that library data should be as open as possible to promote the open science goals. However, the respondent at HEIs libraries indicated their customer's data have never been disclosed. As Person C and E stated data should be "*as open as possible as closed as necessary*". The statement matched with the open science goal . In addition, librarians admitted that HEIs libraries are publicly funded so it is fundamental to share data with the public. They added that the metadata of research and publications should also be open to the public to promote transparency of HEIs.

Person G (ICT) stated that policies, manuals, and guidelines had been shared between the ICT and other units. Moreover, the expert indicated that the major goal of their unit is to safeguard other people's data rather than producing and sharing. the expert indicated that data pertaining to scripts can be reused to understand and improve the existing system.

According to experts from libraries at the HEIs , library and publication data were shared internally with staff and students at HEIs and externally with national repositories such as VIRTA.fi, RESEARCH.FI, and other HEIs repositories as evidenced by the following quotation.

> *"For example, the library data about books, we are already sharing because the program we have is about to share. So, we are open. If there is already a post for the books, we can see and we collect it. Then we put exact items in our system. So, we are already sharing this library data. We also share our publication data. Because you can search the publication search with a DOI (Digital Object Identifier) in the system. So that is why you can download the publications. So, they are also shared and reused with other institutions because researchers sometimes write together with staff from other higher education institutions. If it is reported then, somebody else can of course download the data and use it again. You have to put some data there, but you have to place the base and also this data that's here in* **research.fi**. *We are supposed to be able to download order from there, but it also there is a place called VIRTA where they're actually store so that you can get better downloads if you need"*. **Person D , March 2023.**

The HEIs data have been shared internally with staff, students, researchers, and all academic community and externally with other HEIs and national repositories. The result of the study also indicates that the academic community can also reuse open data for their

activities. Moreover, administrative experts revealed that the GDPR has significant impact on data-sharing activities.

4.2.6 Influence of regulatory environment on business use of data

Respondents were asked question pertaining to the influence of regulatory environment on the business use of data. All participants stated that the regulatory frameworks such as GDPR, national and institutional laws highly influence the business use of data. Participants from HEIs library responded that the GDPR is substantially affects the way they handle personal data.

Person E (library) expressed that the GDPR changed the way they operate due to issues related to cloud services from other counties such as United states of America. The majority of their cloud services are under U.S jurisdiction even if the storage is in Europe. Those cloud services from other countries are governed under their own country jurisdiction which contradicts with GDPR. This causes difficulties when (the) HEIs attempted to use international services. Therefore, the HEIs had to take remedial actions such as security measures before sharing their data regarding their researchers' data and publication metadata and as evidenced by the following quotation.

> *"The GDPR makes it difficult for us. For instance, if we want to put publication metadata into the information system that uses cloud services that are under US jurisdiction, and this we can do without some sort of remedial actions. We're not allowed to just let that information about our researchers or about their publications run free. But sometimes it's difficult for us also because the publications will be openly available online. The law requires us to be very careful with the data flows. So, it's a conflict there because it will be openly available in two weeks. But we have to have a lot of not so much maybe security measurements in place, but we have to before we take a new service into use, for instance, we have to go through a lot of measures to actually map exactly what sort of data can go up there. So, the regulations affect all our activities a lot".* **Person C, March 2023**.

Person C added that *"the GDPR and the US laws. They don't really work together"* to show the enforcement challenges. The ICT expert also stated that regulations such as GDPR impacts data transfer decision to non-EU countries. As a result, when the institution uses cloud services from non-EU countries, the GDPR needed to be referred.

Therefore, according to the ICT expert, transfer impact assessment (TIA) has been conducted several times to ensure GDPR compliance and address associated risks prior to transferring data to non-EU countries.

All respondents concurred that regulations and laws such as GDPR affects the way they handle personal data. The result of the study revealed majority of the HEIs cloud services under U.S jurisdiction which caused challenges on applicability to GDPR in extraterritorial scope. To overcome such challenge, HEIs education had been developing remedial action.

4.2.7 Challenges affecting data value

Participants were asked questions regarding challenges that can undermine the value of data as an asset. For administrative staff most prevalent challenge is related to the GDPR issues. Person A stated that *"I think this GDPR is too strict".* The respondent explained that it has been creating barriers when it comes to accessing personal information to organize institutional events. For instance, after staff or students provided their diet preferences for particular event, that data must be deleted after the event. This is causing redundancy of deletion regarding staff data after they organize events. In line with this, whenever administrative staff want request personal data from academic community, they are required to incorporate policy statements so that concerned individuals can give their consent to it. This makes the processes complicated.

For Person B, the challenge that affects the value of data is digital data itself. According to this respondent *"digital data can be changed".* The respondent added that *"now when everything is digital , it moves, and it is alive. That is dangerous".*

From Librarians perspectives the value of data was constrained by several reasons, namely, lack of expertise, system outage, and lack of data quality(i.e., inaccuracy of data) as it can be appreciated from following quote:

> *"We are very interested in data now and we experience a lack of expertise within our organization. We have a few people who know it, but they need to know more about the organization in order to help those co-workers who deliver services and working to help them know how they can. How to create the data flow that can*

actually benefit us in our both in planning, but also the way resources are managed. Where do we put in more people and money? Where is there a need for development? So, I hope that we can use data much more. And this for us, data is not It's an asset that's actually worth money because we can use it to do things smarter. So very important and. I feel that we don't have known what we need, and we have working on trying to find out what we need to collect and how we need to store it and show it. So, both store it, archive it, retrieve it". **Person E, March 2023**

Person E also stated that besides the limitations of skills and capabilities to generate value from data assets of HEIs, the data flow stops moving if data sensitive which in turn impacts the quality of the data. The format of data, for instance, if it is in the non-standard formats, it may also impact the quality of data which in turn inhibits data sharing and reusing. Person F also added that the absence of analytical tools and competencies hindered to tracking high demanded skills, comparing students' performance in a specific situation such as pre-pandemic and post-pandemic, evaluating students learning preferences (online, offline and, hybrid) and identifying undesirable courses in the market. Therefore, Person F suggested that data stewards or skilled mediators could fill such kinds of gaps.

Person C (Library- research services) stated the GDPR and research ethics have an impact if the study conducted by the researchers involves human beings. Person C added that opening data from other field of studies may also be influenced by intellectual property issues According to the same respondent opening data from medical, biomedical, and clinical research are impacted by the GDPR. Therefore, the finding indicated that the business of uses of data, data sharing, data storing, and data reusing had been impacted by the GDPR.

Person G (ICT) said that the potential challenge which undermines the value of data is data leak which could lead to data monetization by other parties as evidenced by the following quote.

"Well, I see challenges from the economic aspect in general. Research data is valuable, and the data should not leak out. It should be handled in such a way that the researcher and the research group is able to publish the data before it leaks out somewhere else or somebody else publishes it. We haven't seen it, as I know, in our university, but it has happened elsewhere. So, I mean, this is a real sort of this is a

real risk, if you don't handle your data. For me the same, of course, goes for all personal data also. I mean, in H.R., in finance when they're paying wages and of course also, they IT systems for management. Even if there is data that we don't usually use and try to monetize on it, then losing it will not such diminish the value of the data. However, it would be large consequences for us in order to sort of repair the damage of having leaked this data. In research data, if data is released before the researcher intended to release it, then someone else can monetize it". **Person G, March 2023**

In general, the value of data at HEIs are constrained by several reasons. These are lack of expertise, failures in the systems, lack of data quality, data leak, non-standard data format, sensitivity of data, the GDPR restrictions, and research ethics issues.

4.3 Data Quality

According to DAMA(2017), data quality is the central reason for establishing enterprise-wide data governance. Data quality decision domain involves developing data quality measures, policies, and guidelines, utilizing data quality tools, establishing guidelines and plans for data monitoring and cleaning, conducting data quality audits, and identifying data quality challenges. Subsequently, these data quality elements are briefly discussed below.

4.3.1 Policies and guidelines pertaining to data quality

Participants were asked whether there are policies and procedures in place that ensure data quality attributes such as accuracy, completeness, timeliness, and consistency so that stakeholders at HEIs can obtain a good-quality data .

Participants from the administrative confirmed that there are policies and guidelines to safeguard data quality. Person B stated that *"there are written regulatory documents that we have to follow"* around the data quality. However, there are challenges to maintain data quality due to human errors and data systems complexity. They also stated that they perform cross checking on the data to determine its accuracy.

> *"I think that's when you request some reports from the data systems. Maybe you can do cross checking like that. And then, of course, these are very good tools because usually I can see this can't be true, it can't be like this. And then you start to dig. What's wrong? So, all the people who have a very long experience of the job and so on. So, they are good tools too".* **Person B, March 2023**

Person F said that they were following the processes and policies that the whole university is following. As mentioned by the same respondent, *"if any faulty data is detected so that shall be corrected. I guess no data is perfect"*. Thus, possible inconsistencies that affect data quality were corrected in their unit. Furthermore, external auditing has been conducted every five or six years which in turn impacts the data quality of the universities. In other words, external standards have an influence on the processes of data quality.

All experts at HEIs libraries stated that there are guidelines and policies pertaining to cataloging system to maintain data quality. In addition, they affirmed that open science policy recommends researchers to maintain the quality of research data as evidenced by the following statement.

> *"For instance, we have open science policy . It stated that the researchers need to document their data and create meta data to ensure the quality of the datasets. And in practice it would also mean that researchers need to improve the quality by following good principles for data management and the basics of data management"*. **Person C, March 2023**

Person C argued to that open science policy, however, does not clearly state about data quality attributes such as accuracy, timeliness, and relevance in detail. Rather it describes data quality from the aspect of FAIR principle to *"make the data machine-readable"*. The respondent also stated the relevant attribute of data quality can be affected *"if you have funding from funders which demand researcher to deliver research data on specific date"*.

The selected HEIs in this study share their printed materials and e-resources of the libraries between each other. Also, their customer base for printed materials is similar. However, the e-resources are governed individually. Problems related to data quality occasionally occurs if the links to access the e-resources of the other institutions are broken. This in turn affects the quality of the metadata to access e-resources.

ICT expert added that , even though there is no written policy concerning data quality to ensure data quality attributes, the ICT staff at the HEI had been following *"system management"* and *"carefulness"* principles to maintain data quality. The same respondent stated that all ICT staff are "extremely" familiar with data quality attributes.

Participants from HEIs libraries and administration stated that there are policies, procedures and guidelines in place to ensure data quality. ICT expert stated although there is no written policy regarding data quality in the unit, they have been using system management principles to preserve data quality. On the other hand, Person C (Library-Research Service) revealed that the existing Open science policy was inadequate to provide detailed information on maintaining data quality attributes.

4.3.2 Data quality tools and plans

Most respondents considered the question related to the availability of data quality tools and plans either as difficult or more technical. Person F (administrative staff) said that their unit does not own specific tools for monitoring data quality rather they use the existing system the whole university was using.

According to the librarians, the HEIs libraries were using archiving instructions and information management plans rather data quality of tools and plans. Person C (library) expressed data quality tool and plan required technical solutions. On the other hand, Person E stated that the document management system of the HEI has in-built data quality tools, but they do not have such things in their library.

Person D (Library) stated that their academic library does not have data quality tools and plans other than checking their data assets from time to time manually. All librarians agreed that their library e-resources had been checked annually (whether the links to the documents are working). This indicated that HEIs Libraries lack data quality tools and plans to monitor the data quality of the institution.

Person G (ICT) stated that since the goal of the ICT department to safeguard users' data they do not necessarily use data quality tools or plans. The respondent argued that the users' data changes rapidly , and the users have full control over their data they store and backup users' data as long as the user are associated with the institutions.

Majority participants revealed that their units lack data quality tools and plans. The result of the study indicates that document management system at HEIs have data quality tools. Librarians mostly utilize information management plans and archiving instructions to manage their information and library resources rather than their data assets.

4.3.3 Data cleaning and monitoring

Person B (administration) said that there is a designated person who performs data cleaning at university level. The pre-processed data have been used to conduct analysis using tools such as Power BI. Accordingly, the analyzed data can be utilized to generate reports to obtain funds from the Finnish government since 50-53% percent of HEIs budget comes from the government. In other words, HEIs need to negotiate with the government on specific terms such as producing certain level of educated people yearly to gain financial assistance. This indicates data assets at HEIs can be monetized to bring economic value.

All respondents from the HEIs library stated that they monitor the collection and e-resources of their libraries. Experts from the HEIs libraries stated data cleaning may possibly be performed by the ICT department.

Person D emphasized that the e-resources of the library were purchased from vendors as a package. So that they do not have the authority to clean the data they have. Moreover, Person C stated data cleaning is something that research groups often do. However, they provide advice pertaining to tools that could be utilized for data cleaning.

Person E stated that *"I know that the library system's database can become messy and, but I don't think it's not regularly checked to see how it goes. I think it is more like sometimes in between. It's cleaned up a little bit"*.

Person G (ICT) stated that when users of the HEI systems leave the organization their data are stored *"for a certain amount of time"* for verification after that *"their data is wiped"*. The same respondent stated that there are policies and principles in place regarding data cleaning.

In general, all respondents concurred that to some extent data cleaning, and monitoring have been performed in administrative, libraries and IT areas of HEIs. However, Librarians face difficulties pertaining to performing data cleaning on their e-resources because e-resources are purchased from vendors as package which makes it difficult to modify the data.

4.3.4 Data quality audits

According to DAMA(2017), data audit involves assessment data assets of organizations to fit for specific purpose. It often involves developing comprehensive checklist and quality control criteria.

Person A stated that data quality is had been audited to assess whether the existing data is fits to specific purposes. The respondent stated that before reports are generated and sent to concerning bodies, the quality of data had been checked whether it corresponds to the specific purpose. In line with this, Person B said that their unit has been conducting systematic follow-up to strengthen the data quality.

From librarians' perspective data quality issues were addressed indirectly either by checking the issues of datasets or by external organizations such as national publication repositories. As Person D (library) indicated data quality audits were not directly performed in their unit rather they assess the misuse of their libraries data assets. Person C asserted that national repositories , where publications of HEIs are opened for public, may have such kind quality control criteria. In general, at the HEIs libraries there is no indication of utilization of data audits checklist or data quality control criteria.

Questions related to data profiling is not addressed by majority of participants. Person G (ICT) firmly asserted that they do not profile users' data rather they might draw graphs pertaining to the amount of data stored on servers or home directories in general (i.e., not per user)

To sum up, data quality audits at HEIs have been directly through cross checking and systematic follow-up or indirectly by evaluating the misuse of their data assets. The notion of data profiling was not addressed by participants because the regulation limits to conduct such activities since it might violate the laws. On the other hand, the ICT expert

expressed that data profiling related activity, such as measuring the amount of data storage, has been conducted in their unit.

4.3.5 Challenges related to data quality

DAMA(2017) asserts that due to poor quality of data the organizations could lose their profits, reputations, and customer base. Regarding data quality challenges, administrative staff stated that data quality challenges often emerge from people, subjectivity of data, and lack of accuracy.

Person A (administration) stated that "*the challenge related data quality is that is not true., that it is false. That is the biggest challenge. Some data is false and then we have to do something different*" to reveal how impact of data inaccuracy. On the other hand, Person B (administration) said that poor quality of data is caused by subjectivity of the data and personnel issues.

> "*The challenge is always people. Data is not ever 100% safe. For example, the comparison we perform regarding how many registered to courses, how many passed with what grades etc. That is only true only if the teachers are the same. If they're two different teachers. It is not reliable data to compare. Because of its subjectivity*". **Person B, March 2023**

Person F (administration) also concurred that system users mistake can cause threat to data quality said that "*there are people involved in the system. So, there can be human errors. Even though the systems are rather user friendly on such things, I mean to provide correct data*".

The major challenges faced by HEIs librarians arises from researchers, lack of competencies, absence of policies and standards, poor leadership, and lack of trust in employees, and data conversion problem, digital data technical issues. According to Person D, the researcher who reports publication may fail to provide correct data into the library system which in turn causes poor quality of data. As evidenced by Person E (library), capacity gaps and mismanagement could also affect the quality of data.

> "*Well, the challenges arise from lack of knowledge, skills and policy. And there is no clear policy. Also, the collection points and modes of collection, lack of*

standards, bad leadership and the employees need to have secure employment at least". **Person E, March 2023**

Besides, data quality issues were also associated with research data format. Regarding research data, data quality challenges often occurred during conversion of data into another format. Due to GDPR issues video and audio research data need to be converted or transcribed into texts to be opened for the public as shown by the following quotation.

"Well, when you transfer data or if you convert data, there might be some challenges there. For example, in the transcription of the interview data from audio files to text, there are some challenges where that can be mistakes and it can be checked by another researcher to ensure the quality. That is something that we advise in when they write their data management plans and how they should manage these challenges. And when you transfer data or when you digitize data, there can be technical issues or other challenges". **Person C, March 2023**

Person G (library) indicated potential challenges pertaining to data quality are associated with research data storage. In other word, if researchers store research data in their personal space without adhering to data management plans, it can cause the risk of losing data or data leak as evidenced by the following quote:

"If researchers save data on their personal storage media, these are the problem for the data quality, because when they leave, we don't have control of what data they have there. So, we will wipe the data and the access to the data they have had in there. Their cloud storage also goes away in a certain amount of time. Hence, if they have not sort of adhered to recommendations about how to store data, then that would problems. So, the biggest challenge is not storing data in accordance with the data management plan". **Person G, March 2023**

In addition, risk of ransomware encryption was another threat to data quality. However according to ICT expert data securities risks such as ransomware encryption had been handled by the ICT department through regular backups.

As mentioned above, multiple factors can hamper data quality at the HEIs. These were lack of standards and policy, absence of skills and knowledge to manage data, inaccuracy of data, human errors, poor leadership, subjectivity of data, ransomware attacks, failure to adhere to research data management processes and recommended data storage solutions, data conversion issues, and lack of technology.

4.4 Metadata

DAMA(2017) asserts that metadata is essential to manage data quality because it clarifies what the data itself represents. In other word, since the aim of data quality decision domain is determining requirement of data use intentions, well organized metadata have the capacity to improve the quality off data.

According to administrative to staff' response, metadata is important to make student-related decisions (e.g., students' admission). However, there is not specific plan to document semantics of data, the use and storage of metadata is determined by policies related to individual rights. The respondents mentioned that sensitive data such as student or staff medical meta data are required to be discarded after specific decisions. Thus, the storage of metadata is determined by content of the metadata itself. Other type of metadata that administration staff utilize frequently is study metadata which is often stored for longer period for verification purposes. Thus, it can be determined that most of metadata used for decision-making by the selected HEIs is domain-specific metadata. The physical metadata in the selected HEIs are modelled and updated by third parties who had an agreement to provide learning management system, students data system and document/ record management system. As indicated by Person B, the document management system supports recording, sharing, and storing data. The system contains metadata related to type of decision made, the authority who made the decision, data ownership and access rights in which data owners were able to make changes or updates. Additionally, Person F indicated that the document management system has in built system to categorize(documents) and tag data(metadata). Therefore, the system helps users to define, and utilize the metadata assets of the institution.

Person G (ICT) described that metadata were mostly used in research data management for sharing and reusing of research data. The respondent indicated that the ICT department only uses quite limited metadata for decision making.

Among other areas, metadata appears to be essential for librarians. All participants from HEIs libraries confirmed that there are guidelines and procedures that help library system users (students, researchers, and library personnel) to create metadata. All respondents at

HEIs agreed that researchers are responsible for creating their publication metadata, including metadata of their datasets, using the library systems or the national databases.

Person D described that metadata allows users or the public to identify what the publications are about. The HEIs libraries collect publication metadata(i.e., books, articles, and manuscripts) from national repositories and share their metadata with national repositories and other HEIs. According to the same respondent, metadata were also used to create parallel publications. In other words, publications can be accessed both from national repositories and HEIs libraries.

Person D also stated that there are certain procedures, and standard (e.g., Research Data Alliance) that researchers are required to follow to create metadata before they open their publication or research data for the public. The respondent added that metadata of publications are defined based on the guidelines and forms provided by the HEIs library system and national repositories. Therefore, librarians are responsible for authenticating and publishing the metadata after approving that all mandatory fields have been filled out. Person E (Library) also concurred that there are guideline and standard such as Research Data Alliance that help researchers and librarians create a good metadata in relation to research and catalogues by saying that "w*ell, there are for research data. There are some guidelines on how to create good meta data for research data. And also, for librarians and cataloguing people. They use RDA standards. The Research Data Alliance*".

Person C (Library- research services) added that they recommend researchers to use the FAIR data services offered by company such as CSC (IT Center for Science) to structure and define the metadata of their research datasets. Consequently, the metadata can be published on other research data services such as ETSIN (metadata finder). These were evidenced by the subsequent quote.

> *"Well, basically, we recommend the Fair Data services that is offered by CSC. And then when you describe your metadata in the servicee, you get a structure for what is required, you fill in every step and then then you basically say everything you need to say about your data sets. And then this data, this metadata can be then published in another service, which is called ETSIN, which is a finder for datasets in Finland".* **Person C, March 2023**

Besides, Person C stated that they also recommend researchers to store their datasets internally to the university's research information system. However, they claimed that majority of researchers at HEI had not been publishing their datasets. Therefore, to provide additional research materials for the academic community, they considered to add metadata of research datasets from external repositories to their research information system. However, the existing tools to perform such tasks are unsatisfactory and they are still under development.

To sum up, metadata have been utilized for different activities and decision at HEIs. Systems such as document management system, student data systems and national study application systems were supporting to generate and access study, student and staff related metadata later utilized for decision making. Among all areas metadata have been effectively used to support the HEIs libraries activities. For instance, research metadata have been used to create parallel publication. In line with this, HEIs were sharing their library collections, research datasets, and publications metadata with other HEIS and national repositories to improve accessibility of resources by the public and research community. On the other hand, researchers at HEIs fail to publish their metadata internally often. Thus, the HEIs were forced to use external repositories to collect research datasets metadata. However, the existing systems at HEIs libraries were inadequate to collect metadata of research datasets from other external repositories. As a result, the research community may face difficulties to acquire rich datasets through HEIs libraries.

4.4.1 Metadata Modelling and Updating

Metadata modelling and updating enhance data users to obtain data from different data sources (DAMA International, 2017). Person E(library) stated that to keep the library cataloguing system up-to-date and to model library metadata of the institution, they were using MARC (MAchine-Readable Cataloging). According to the same participant their metadata flows follow international standard such as MARC to exchange and interpret bibliographic data. Also, the respondent indicated that currently there is an undergoing national project that targeted at shifting MARC cataloging standard into another cataloging standard called bibframe.

"Well, the library system now using the MARC format. And it will be changed to bibframe. And this is a national project.... MARC it is XML. Yes. And it will be changed to bibframe. So, this is a project for the libraries that were involved in this metadata Data exchange". **Person E, March 2023**

Person C (library) stated the issues of data (metadata) modelling have been discussed on the national level to solve communication problem between various repositories in Finland. The discussion is still ongoing to find technical solution to solve communication problem. Besides, person C added that the national-level discussion also incorporated the possibility of making more structured data management plan at HEIs in the future. Person C argued that keeping research metadata up to date is difficult task because only researchers can modify their metadata or datasets. The respondent claimed that there are no solutions from the personnel side for such issue as evidenced by the following quote:

"I think for updating the data that there are no solutions for that. I would say that is something that is just the work of the researchers.... Perhaps, creating metadata is something that researchers often do at the end of the project. But when you write a data management plan, you already think about these issues. In the beginning of the project. So, I think it could be wise to increase the awareness of creating meta data throughout the research data process and the whole life cycle". **Person C, March 2023.**

Person G (ICT) also supported Persons C's argument saying that "*I can not see a reason to update the metadata unless the data behind it actually changes.*" Person G indicated that researchers, researcher groups, and data owner are in control of changing the metadata if needed.

In general, standards such as MARC (MAchine-Readable Cataloging) have been used by the HEIs libraries to model and update the library resources. Research related metadata were difficult to update because only researchers had the authority to modify research data.

4.4.2 Data dictionary

According to DAMA (2017), data dictionaries are useful to justify data from the aspect of business terms so that all stakeholders can understand about the data. All respondents

from administration stated that each unit in administration have different data dictionaries.

According to Person B (administrative expert) for each system related to students such as education services, student data system, and the national study application system, there are data definitions, descriptions, and dictionary. Particularly, Person B said that *"we created a dictionary and handbook around that student data system we developed. It is still not ready; it's still being updated and developed"*. The document management systems at HEIs also have manuals that guides staff to become familiar with terms related to the systems. Person F (administration) stated metadata regarding course participants were not fully utilized due regulatory and ethical issues.

All respondents from HEIs libraries stated the research data guide provides detailed instructions and definition of key terms which in turn can be considered as data dictionary or business glossary. Data dictionary strengthens the understanding and communication between librarians and researchers. Besides, librarians stated that it could also serve as standard for metadata interpretation. Person E(library) revealed that their institution has data dictionaries in place to understand bibliographic metadata as evidenced by the following quote.

> *"I think, for instance, for the bibliographic data…I think there are data dictionaries. And I think they are crucial because they make sure that everyone understands the same thing and that they are exchangeable and standard for it. And also, for the MARC of these data models".* **Person E, March 2023**

Person C (library-research services) stated that the issues related to data dictionary were discussed in both research data management plan and research data guide which in turn facilitates the communication between librarians and researchers. Specifically, the research data guide provides detailed instruction for creating data management plan and defining important terms that researchers are required to familiarize themselves with. Besides, the guide offer quite detailed instruction in relation to GDPR issues that researcher, teachers, and students need to know before they publish their research data and scholarly works. Person (ICT) stated that their unit has been using the National Data

Dictionaries to build common understanding on the terms related to data and cyber security.

All respondents indicated that their institutions have data dictionaries in place to communicate with stakeholder from research community, staff, students, and government bodies. However, metadata related to students' personal data have never been used due to ethical and regulatory issues.

4.4.3 Challenges pertaining to metadata

Person D from library stated that accessing publication metadata from national databases such as Theseus (a repository for Finnish Universities of Applied Sciences) can be constrained by IPR (Intellectual Property Right) issues. In other word, if the publication is licensed, it is difficult to access without the publisher or researcher authorization. As a result, it hinders the HEIs libraries to open publications parallelly. In line with this, the existence of temporary links to the publication could also affect the quality of metadata because it impedes users to access the publications they are interested in when they need it. However, parallel publications created by the HEIs is aimed at solving such problems to get permanent address to the publications.

According to Person D (library), there were occurrences when researchers failed to provide necessary information about their publication. Thus, the librarian had to follow-up publications made by their researchers by themselves, manually, and report how many publications they produced during the year and in what field of study they undertake their research. So, this has been major challenge for the librarian for the past years.

In general, challenges regarding metadata stemmed from carelessness of researchers to publish metadata, temporariness of (links) metadata to access publication, and IPR issues.

4.5 Data Access

The aim of data access decision domain is to determine data access for organization. In this section, data access elements such as ppresence of policies, procedures, standards, backup and recovery program, mechanism to disseminate security education, risk assessment plan, and regulatory compliance at HEIs are discussed.

4.5.1 Data Access Policies and procedures

Respondents from administration stated that for each section under administration such as human resource, student affairs, finance, and faculty administration, there are certain policies and guidelines that limits data access. Besides, those policies and guidelines are developed based on national laws and regulations as illustrated in the following quotation:

> *"Yes, we have policies and procedures. We have for each of the sections such as Students services, HR, and so on. We have restrictions. We have documents that the tells us how we shall handle data. We also follow the national level laws for example, the GDPR".* **Person A, March 2023**

Person B stated that each section can only access data that is concerned with their tasks. Thus, data access is directly linked to the role of users as evidenced by the following quote.

> *"For example, in student data system, there are different kinds of roles. Someone is teacher, someone is a planner. So, I think you can see different data for doing different things. And also, if you don't work with admission, you don not have access to the national system such as the studyinfo . So, you have to have permission to gain access".* **Person B March 2023**

Person D (administration) also concurred the policies and procedures regarding data access in their institution were defined according to the role and level of each user group as evidenced by the following quote:

> *"We are using the same policy that the University is using. So, there is some data that's open for all internally. Then some data that's open for the whole internet. And then there is some data that is only open for staff members. And there you have certain levels of access. For example, the top management get access to specific issues. In Finland there are quite strict data security rules for certain types of data like such as health data, and personal data. Accessing such data is very limited. In overall, we are following the same policies and procedures that that the whole university is using".* **Person F, March 2023**

All respondents at the HEIs libraries said that there are policies and procedures that restrict and monitor data access. These policies also consist of data storage and security solutions.

Person C (library-research services) stated that existing guidelines specify the required data storage and security solution pertaining to their data assets. The same respondents indicated that supervisors at the HEIs are responsible for providing advice to students regarding research data storage and access.

Person D (library) revealed that most of the resources managed under their library are "*open as much possible*" as they intended to avoid making differences between people as shown by the following quote :

> *"For the resources we are managing through our system , there are no restrictions. Everybody who has the university email and passwords can access it.So, there should not be any restriction on that. But there might be restrictions that only a few people can use it at the same time. But that's something we are working on it, but we don't have how restrictions on different kind of staff, as I far as I know".* **Person D, March 2023**

As mentioned above by Person D, there are possible restrictions in relation to allowing multiple users to access the same material at the same time. The same respondent added that there are also some sensitive materials from medical and maritime department that could only be accessed by the staff otherwise most of the resources opened for anyone with access rights. The respondent revealed that, nowadays, there several restricted collections in the national databases that require access grants to read or download the publications.

Person E (Library) stated there are certain instructions within the institution that restrict different user groups data access based on level of employment and responsibilities. However, there is no written policy regarding data access rather issues of data access had been addressed in the task descriptions as evidenced by Person E:

> *"For instance, in our library system accessing and we have user groups, so not everyone can access the data about the users. For instance, we have to the people who own the books, the customers, we have to the people who help them. But then we have many different modules in the library system and not everyone can access everything. But we don not have written policy, I think, it is included in other documents such as the task description. It follows along with what your work is about".* **Person E, March 2023**

Person G (ICT) stated, there are certain standards in place which authorize users' data access right. The same respondent indicated that every user has its own personal account to access different systems and data (i.e., depending on their roles at the institution).

Majority of the respondents stated that there are policies, standards, and procedures that restrict data access at HEIs based on their role and status. In few cases, as a replacement for policies and procedures, data access restrictions were reflected through task descriptions and instructions. Although majority of library resources are opened for data users who have access rights, there are specific restrictions regarding certain publication that requires data owners' authorization.

4.5.2 Consistency of policies and procedures with GDPR

Participants were asked about the consistency of policies and guideline with the GDPR. Administration staff responded that the existing procedures and policies regarding data access are consistent with GDPR. Person B (administration) mentioned that even though there is non-disclosure agreement document to share data among the administrative staff, every staff understands the GDPR since the university and national laws states that people who work in educational institutions must be cautious or comply with it, to maintain data integrity. Person F (administration) also believed that their unit has been functioning according to the GDPR rules since the systems, guidelines and procedures at the HEI were built in accordance with regulation (GDPR).

All respondents from HEIs library confirmed that the existing guidelines are consistent with laws and regulations pertaining to data processing. Person C(library) and Person E (library) stated that the existing data access guidelines are consistent with regulatory frameworks such as GDPR as evidenced by Person C saying that *"yes, they are shaped in relation to GDPR."* Person E also agreed by saying " *I think they are consistent with GDPR. Yes, I think at least it's minimized".*

Person D (library) also added that their customer data are not allowed to be read by anyone else they grant access to that material or data. Besides, only users who have access right at HEIs are allowed to access the databases. Person G (ICT) stated that all the data handled in their institutions are GDPR screened so that users don not have access to data

they were not supposed to use or access. All the respondents concurred that policies, procedures, and instructions related to data access comply with the GDPR.

4.5.3 Internal procedural controls

Data access decision domain also consists of decision related to data storage. According to administration staff, European Union (EU) related documents should be stored for five years for auditing purposes. The respondent indicated that there are different plans for different types of documents with respect of their storage and deletion time as evidenced by Person F emphasizing that *"for Certain EU projects you need to store every single piece of documentation and the data for certain number of years"*.

In addition, Person A revealed that internal procedures have been established by legal office (data protection function) of the university to handle user access issues. The respondent said that data protection function at institution is responsible for setting internal procedures regarding data processing and data access.

Person F (administration) stated that there are *"ready-made agreements"* that have been used to make a deal with project partners due to the IPR issues. Consequently, for project related to continuous education, there certain procedures that define who own, use and utilize the data or result of the project. The respondents also suggested that it could be beneficial for their unit if they begin using non-disclosure agreements.

Person C (Library) from research services stated the legal and ethical issues should be considered before opening research data. The respondent emphasized that "so that is why it's often recommended as open as possible and as closed as necessary." In addition to that Person C revealed that the open science team at the university also provides advice about data management pertaining to how to handle sensitive and confidential research data in the best way.

According to Person E (ICT), ICT department had executed security measures such as Multi-Factor Authentication (MFA) to monitor users' data access. In addition to that, all respondents concurred that the Finnish Ministry of Education and Culture funded CSC to develop data access standards such as HAKA login and eduGAIN to safeguard users' data access at HEIs.

In general, to manage users' data access at the HEIs, standards and security measure such as Multi-Factor Authentication (MFA), HAKA login, eduGAIN, and internal procedures (by the HEIs data protection function) have been established.

4.5.4 Backup and recovery programs

To prevent data from potential loss, backup and recovery programs or plans are necessary for HEIs. Administrative staff explained that backups, recovery , and risk assessment plans (i.e., Data security)are conducted by ICT department. Yet, they are aware that the ICT department backup copies every night on the servers. They also stated that there were few occasions when systems such as student data system was down. However, the ICT department took corrective measure to restore the system.

Person G (ICT) stated their unit have clear backup and recovery program that safeguards institutional data assets. In addition, ICT expert stated ICT staff conduct risk assessment regularly to identify potential threats and risks to the systems. The following quote seconds that the ICT department in charge of backup program at the university level by mentioning that *"that is something that our IT department do. So, they have a backup for our database. I think they take it everyday. Of course, they did delete the older ones, but I think they have more than one backup"*.

Pertaining to data storage solution, Person C said that the most reliable data storage solution for research data is to store them at the HEI servers as indicated by the following quote:

> *"And mostly the most secure solutions for research data are to store them at the university servers, which are backups regularly, and they are also, mostly it is the best solutions for sensitive data. And if you only, um, if you are one person in the project, then the best way to store the data is to at home a storage area at the university service. We think this is very safe"*. **Person C March 2023**

Person C stated that the research data management team offers specific advice about how long data should be kept for verification purposes. Moreover, the decision regarding research data storage and verification period varies between disciplines. The same respondent highlighted that the storage solutions at HEI would become graveyard for old

datasets as there was no policy that addresses such issues. Furthermore, Person C revealed that "*when people no longer are affiliated to the university, they will lose their storage capacity at the university, which means they should transfer data to something else or keep their affiliation. Yes. Otherwise, you might lose your data*". Therefore, absence of clear policy related to research data storage solutions has an impact on both researchers and the datasets itself.

As mentioned by Person C (library), personal data are treated differently because the laws and regulations such as GDPR affect the data storage plan. For instance, audio files from interviews should only be kept for a limited time and then deleted.

Person E (library) stated, data pertaining library user is only kept if "*there is a grounds processing and storing the data*". In other words, Person explained that "*as long as someone is a lender, the account needs to be available*". The same respondent indicted that the GDPR enforces that there should be a ground or basis for processing users' data.

Person D and Person E stated that the ICT department is responsible for decision related to storage solution (i.e., for how long data should be kept) at HEIs. Whereas national repositories developed by other external parties, who provide data storage solution, are responsible for planning backup and recovery program as evidenced by the following quote.

> "*I would say that the company CSC, which is monitoring this business. I would say they have. They do their best to backup copies because there has never been a question that we should do it ourselves. So, I suppose they have backups*". **Person D, March 2023**

Person D added that librarians worry that if internet down, they might lose data pertaining to loaned books since the library system is digitalized. However, since most the publication have been published parallelly with national repositories , Person D stated that they were less concerned regarding this issue.

In summary, backup and recovery programs concerning internal data sources and publications made by the HEIs academic community had been handled by the ICT departments at HEIs. Whereas backups regarding external research data had been handled

by national repositories. Additionally, the respondents indicate that regulations such as GDPR have an impact on data storage plan.

4.5.5 Dissemination of security awareness and education

The participants were asked questions pertaining to how security awareness and education is disseminated in their institutions. Administrative experts responded that education and awareness of security measure were disseminated through different ways such as workshops, mandatory information security course, emails notifications (i.e., regarding phishing attacks) and individual solutions. Person A (administration) indicated from the aspect of the GDPR *"when the GDPR came, we had a lot of workshops and things like that. So, we tried to deliver the information to the people"*. Person B (administration) mentioned the mandatory information security course given at their institution by saying *"for example, for students, there is mandatory information security session. Everybody has to complete the course. Also, the university gives individual solutions, if you have some kind of diagnosis"*.

In contrary, Person B had doubts that if the users are unable to complete the mandatory course at the given time frame, they might lose access to their data. According to person B retired professors or uninformed student might lose their data permanently. Thus, this may create a challenge for such kind of massive operations.

All respondent from the HEIs library confirmed that ICT department is responsible for creating security awareness at institutional level. Besides, all respondents concurred that the mandatory information security course offered to everyone who is affiliated with HEIs is also another mechanism to create the awareness about security risks. In addition, Person C specified that information security manager at the university is responsible for creating security awareness at the institutions. The ICT expert agreed that besides courses, emails and workshops, their unit also uses the intranet of their institution to disseminate data security education and instructions.

According to Person C confidentiality issues could pose a threat to data access program of the HEIs. Person C reasoned that when HEIs collaborate with different industries there might be risk of stealing other people or researchers' idea which could cause massive

financial crisis. All respondents from HEIs libraries concurred that, resource, and information retrieval guide (research data guide) and data management plan at HEIs were developed in way that staff and students could comprehend issues related to IPR or confidentiality.

All the respondents mentioned that security awareness at the HEIs have been disseminated through courses, workshops, guidelines, instructions (through intranet)and emails.

4.6 Data Lifecycle

Data lifecycle decision domain concerns with data inventory, data archiving, regulatory compliance (regarding data inventory and archiving), data deletion, and data retention policies, procedures, programs, plans, and activities of organizations.

According to administrative staff' the bare minimum of data is archived. There is also policy concerning this. However, a designated data inventory plan which involves listing of all data assets of the HEI was still missing. Person F (administrative staff) revealed data retention and deletion decisions and rules vary between projects when it comes to project data. On the contrary, data pertaining to training courses delivered by the continuous education unit were stored in public databases since the Finnish ministry of education coordinates and funds the projects.

Person G (ICT) stated that data archiving is done by separate institutions. However, their department is responsible for creating backups. Person G indicated that data archiving could be more important for finance and HR departments for auditing purposes.

Data lifecycle activities at HEIs involves external (national) organizations .Person C (library) stated that research data management team at the institution recommends that data should be archived on external repositories (i.e., national) for long term because the national repositories open the data for unlimited time. According to the same participant, most researchers prefer to keep research data for limited time for verification purposes. In addition, it is possible for researchers to store their research data internally on the servers at the HEIs.

Person C concurred that clear guideline for data archiving are missing. According to Person C and Person E the libraries at the HEIs had physical and digital archives (still under development) to store valuable cultural heritages. Person C emphasized there has been discussion regarding developing policy related to data archiving evidenced by the following quote: "…*well, that is something that we should develop, and we have some discussions about it, but not yet the policy*". On the other hand, all respondents from the library revealed that they have been developing statistics regarding the number of collections their institution owns. However, data inventory has never been done at HEIs libraries. From librarians' perspective, data management plan(DMP) was utilized to manage research data lifecycle. The current DMP at the HEIs addresses questions related to data retention and deletion periods. Person E mentioned that there is information management plan at institutional level that could guide data lifecycle issues. However, Ladley (2012), argues that information management program emphasizes on managing organizational information assets rather data assets.

As person C revealed, majority of researchers at the HEI write their data management plan at the beginning of their research because it is required by the funder. The same respondent indicated that they have been offering advice to researchers to update their DMP at later stage of their research as indicated by the following quote:

> *"No, not really A program. I would say the data management plan is a tool for that. There are questions in the data management plan also about , for how long will you keep data and what parts will you delete? And this is something that should be updated through the research data lifecycle. But many people only write it in the beginning because the funders require it. And we always try to remember and remind people this should be updated, and it can be used as a tool for".* **Person C, March 2023**

Person D indicated that are certain restrictions regarding data retention and data archiving pertaining to library data . Person E stated that the library data is archived through the existing library system as evidenced by the following quote.

> *"Our data in the library system is archived there in the system. The data are exported and archived; we have lots of old catalogs and things like that. They were analogue before. Now, they all are in the integrated library system. So, this would be library data".* **Person E, March 2023**

Librarians added that the Libraries at the HEIs had been inventorying their library collection rather than their data assets. However, they recommended that the digital data assets at the libraries need be known and inventoried. According to Person E (library), management data/information at the institution have been archived through document management system systematically.

4.6.1 GDPR and Data Lifecycle

The experts from administration added that data retention and archiving policies of HEIs strictly adhere to data protection laws such as GDPR. For instance, handwritten material cannot be stored. Moreover, the data protection office is accountable for creating policy and procedures related to the data retention and archiving. The procedures and policy are still under development. The finding indicated that data management personnel, lawyers and information specialists are working together to develop the document.

Person C (Library: research services) said that the GDPR affects data retention, data deletion and data archiving issues because before research data is opened for the public it should be anonymized and pseudonymized. Furthermore, Person C stated that the code list of research datasets should also be deleted to assure that the names of the participants could not be traced. Person C stressed that the library guide at the institution provides information (i.e., links) in relation to IPR and GDPR issues for all staff, researchers and students. Person E (library) also added that the GDPR had impact pertaining to data lifecycle (data production, retention, and deletion) issues. According to this respondent, GDPR influences how they handle their library, research, and personnel data assets as evidenced by the following quotation:

> *"I think it affects a lot. I think that it has affected because when, for instance, the recruitment system deletes the data after three years. So, I do think it affects. And for research data, I think also nowadays the researchers are getting more and more knowledgeable about how they can and should if they can retain or if they should delete, it's their data and that they have permission to share if they want to do that. To make openly available. So maybe not in all fields yet, but in many cases. For also for library data. The personal information that we keep on to users is already handled. It has already been GDPR screened".* **Person E, March 2023**

Majority respondents revealed that there are policies and procedures in place concerning data lifecycle activities such as data retention, data deletion, and data archiving at HEIs. Those policies are affected by and strictly adhere to data protection laws such as GDPR. From librarian aspect the existing guidelines lack detailed information. According to librarians, Data Management Plan (DMP) is essential tool to manage research data lifecycle. The result of the study indicates data retention and archiving are still evolving to fulfil regulatory compliance issues. On the other hand, data inventory plans were missing in all areas at the HEIs.

5 DISCUSSION AND CONCLUSION

This study is aimed at exploring the challenges and opportunities of data governance practices in Finnish Higher Education institutions (HEIs). To achieve the goal of this study relevant data governance framework focuses on decision-making rights and locus of accountability regarding institutional data assets. Therefore, the data governance framework offered by Khatri and Brown (2010) which consists of five interrelated decision domains, namely, data principles, data quality, metadata, data access, and data lifecycle was used to assess significant issues related to data governance practice. Therefore, the key findings of the study are discussed and concluded from the aspect of the research framework, research questions, and literature review.

5.1 Discussion

In this subsection, the research questions of the study were outlined to frame the key results of the study.

RQ1: How are data governance practices carried out at Higher Education Institutions?

Previous studies stated that data governance is aimed at maximizing the value of institutional data assets by organizing and implementing frameworks, structures, roles, policies, guidelines, procedures, accountabilities, and responsibilities (Ladley, 2012; Otto, 2011; DAMA, 2017). In general, data governance focuses on governing institutional data assets and managing data as an asset (DAMA, 2017; Khatri & Brown, 2010). The key findings of the study revealed that in both Finnish Higher Education Institutions (HEIs) data are considered as institutional assets. The HEIs have developed several policies, standards, and guidelines to manage their data asset.

The finding of the study revealed that both institutions do not have data governance separated departments. The majority of data governance activities have been performed through data management strategies of the HEIs. As argued by Ladley (2012), data governance is not about formulating a stand-alone department rather it concerns designing a framework to govern institutional data assets. The author also indicated that in most cases data governance program is devised by IT and business professionals virtually (Ladley , 2012). Multiple data-related activities at both HEIs implicated that data

governance practices have been undertaken. There are various policies and procedures regarding data principles, data quality, management, metadata data, data access, and data lifecycle. For instance, research data in the HEIs is governed by multiple policies and guidelines such as the open science policy and declaration, data management plan, and research data guide (libguide). The HEIs libraries also have employed metadata management standards such as MARC to manage and catalog their metadata assets.

Data governance also calls for the establishment of authorities and responsibilities to manage the data assets of institutions. The finding of the study indicated that for the majority of data-related activities, at both HEIs, different roles and responsibilities had been allocated. In each unit, the role of data owners, and data users were described in the policy documents and task descriptions. Although the role of data stewards was defined in the libraries-research services area which operated under open science and research functions of the HEIs, the results of the study showed that other professionals such as study advisors, finance and human resources, ICT services have been acting as data stewards. The previous study also mentioned that even though organizations lack designated data governance programs or department, most organizations have data stewards who already have been involved in assisting institutions to minimize data-related liabilities, enable accessibility of high-quality data, and obtain value from institutional data assets (DAMA, 2017). This signifies that the HEIs are required to define and formalize the role of data stewards in each unit to unlock value from data assets. Besides, data owners, data users, and data stewards, the HEIs also have incorporated the role of Data Protection Officers who are responsible to oversee personal data processing activities and offer advice on compliance with data protection issues at HEIs.

The finding of the study also indicated that regulations such as the GDPR had an influence on all data governance decision areas as well as the corresponding policies and procedures in the decision domains. In the data principles decision domain, the GDPR influences the business use of data and data sharing because processing personal data regarding students, staff, and external customers requires legal justifications. Moreover, both HEIs have collaborated on different projects and research with non-EU countries. Besides, most cloud services at both HEIs are under US jurisdiction (US-based cloud services). Hence, when HEIs share their research and researchers' data with non-EU countries and use

cloud services from non-EU countries, the GDPR, supplementary laws and rules (IPR, Data Protection Act, University Act, and UAS Act), and remedial actions have been considered to protect personal data collected their institutions. Article 3 of the GDPR also confirms that the regulation applies to non-EU-based companies who offer services to residents of the EU/EEA. The GDPR and IPR laws also greatly impact when HEIs attempted to open research outputs (research data and publication) because some research data consists of sensitive material (medical, biomedical, and clinical research data) and subjects (human beings). Moreover, in such cases the research and The HEIs the open science and research policy and declaration.

The finding of the study revealed that data quality, metadata and data access decision domains, and pertinent policies and procedures were influenced by the GDPR because HEIs have been engaged in openly publishing research data and scholarly work to promote the open science and research agenda. In such cases, the GDPR strictly advocates to anonymization and pseudonymization of research data to maintain the quality of data, privacy research (data) subject, and integrity of the original intent data collections. Furthermore, elements of data lifecycle decision domain such as data archiving, and data retention policies of the institutions were also consistent with the GDPR.

To conclude, at both HEIs most of data governance practices have been done through data management strategies. Policies and procedures pertaining to data governance decision domains were highly impacted by laws and regulations such as GDPR. Moreover, they were developed in accordance with GDPR.

RQ2: Which benefits do higher education institutions obtain from utilizing data governance practices?

As mentioned above, at the HEIs data is considered as institutional assets. The result of the study indicated that administrative staff, librarians, and ICT experts at institutions had gained various benefits in managing data assets through developing and applying policies, procedures, and standards, and delineating pertinent responsibilities and authorities. The finding of the study revealed that data (internal, external, and research data) are crucial assets that HEIs professionals utilized to develop reports and statistics, conduct research, publish research datasets and scholarly works, measure performance,

assess customer needs, provide services to different user groups, devise strategic plans, apply for accreditation, and make strategic decisions. The reports and statistics generated by administrative staff and librarians consist of information such as student-teacher ratio, research output ratio, number of students and researchers in each faculty, number of library collections, and number of students who have graduated and completed their master's thesis in each faculty. The output of the reports and the statistics had been used to make strategic decisions internally (by the management of HEIs) and to evaluate the performance of higher education externally by the Ministry of Education and Culture. The finding of the study was also supported by both Universities' Act 558/2009 Section (51) and Universities of Applied Sciences' Act (932/2014) Section (45) which state that the HEIs are required to provide data necessary for statistics and evaluation when requested by the Finnish Ministry of Education and Culture. Moreover, as the ICT expert revealed, data are extremely significant assets to assess storage allocation and to create awareness regarding data security risks and issues such as phishing, data breaching, and misuse at the HEIs.

The metadata generated through the library systems of the HEIs has been utilized to foster collaboration and cooperation of HEIs with Finnish national repositories and other HEIs. In the same way, the HEIs were collecting metadata of publications and research datasets from national repositories to promote research activities and to overcome the limitations of using data from multiple sources. Thus, researchers can access data from multiple sources for conducting studies which in turn promotes data sharing, data reusing, and innovation. The HEIs have utilized the Research Data Alliance (RDA) standards to create good metadata which enabled data sharing with national repositories and other HEIS. The HEIs also benefited from applying the FAIR principles to research data which in turn augments data reusability with all stakeholders (Wilkinson et al., 2016). In addition, FAIR data services offered by companies such as CSC enhanced the HEIS to structure and define research and publication metadata. Alongside, The HEIs libraries utilized standards such as MARC (MAchine-Readable Cataloging) to structure and model bibliographic metadata in their library cataloging systems. The result of the study indicated that the HEIs libraries are currently involved in a national project to replace the MARC standard with BIBFRAME (Bibliographic Framework) to reduce the cost and time of the traditional cataloging system and maintain data authority.

To promote data sharing and data reusability, and to manage the research data lifecycle in accordance with the national open science policy, HEIs devised multiple documents and tools, namely, data management plan, data management tools, research data guide, and data management guide help researchers. Besides, data stewards at the research services department had the responsibility to give advice to students, staff, and researchers regarding research data management and publishing research data.

The HEIs have been handling institutional metadata flow through different systems such as learning management systems, students' data systems and document/ record management systems, and library systems. The presence of such systems helped the institutions to oversee their metadata.

From a data access decision domain perspective, the establishment of security measures such as Multi-Factor Authentication (MFA), HAKA login, eduGAIN, backup and recovery programs, and internal procedures advanced the HEIs to manage users' data access and minimized data security risks. The finding of the study revealed that Transfer Impact Assessments (TIA) been conducted several times to ensure compliance with GDPR and minimize data privacy risks before transferring personal data to non-EU countries. Furthermore, the existing policies, and guidelines also restrict data access based on the level of employment and responsibilities.

The majority of respondents revealed that there are policies and procedures in place concerning data lifecycle activities such as data retention, data deletion, and data archiving at HEIs. Those policies are affected by and strictly adhere to data protection laws such as GDPR. From librarians' aspect, the existing guidelines lack detailed information. According to librarians, Data Management Plan (DMP) is an essential tool to manage the research data lifecycle. The result of the study indicated data retention and archiving are still evolving to fulfil regulatory compliance issues. On the other hand, data inventory plans were missing in all areas at the HEIs.

The finding of the study revealed that data dictionaries have been utilized by administrators, librarians, and ICT experts to "make sure that everyone understands the same thing." As concurred by DAMA (2017), data dictionaries are essential to create understanding among all stakeholders regarding existing data assets in the form of

business terms. All participants also concurred that the ICT department in each HEIs is responsible for decision-related to the data storage solution

To sum up, the presence of data-related of policies, guidelines, instructions, systems, processes, and standards indicates that the selected HEIs were practicing data governance to manage their data assets.

RQ3: Which obstacles, both existing and potential, prevent data governance practices in higher education institutions?

The findings of the study revealed that several factors can hamper the data governance practices of the HEIs. Major challenges to practicing data governance stemmed from the absence of policy, formalizing, and assigning roles and responsibilities, lack of tools and plans regarding data quality audits and data inventory, etc.

As mentioned above, the role of data steward was noticeably defined in the research services (research data management) area which operates under the Open Science and research team. In other, words, the data steward is only responsible for ensuring the quality, and management of data related to research activities. The role of other experts from IT, Finance, and Human resource who have been acting as data stewards was not officially formalized in HEIs. As reported by KMPG (2018), data stewards play a significant role in ensuring data quality, security, and accuracy throughout the organization. As a result, the absence of defining the role of data stewards can affect the quality of data assets. Alongside this, the result of study showed that librarians, and administrative staff are in need of data stewards.

Moreover, the study also revealed there were gaps in assigning accountable individuals for managing data at the strategic level due lack of data governance frameworks that outline comprehensive accountabilities and responsibilities for data-related activities at both HEIs. This implies that HEIs may suffer from utilizing data to meet their institutional objectives. Previous studies at HEIs in Canada indicated that data trustees are accountable to oversee what happens to university data at a strategic level (Plaid Consulting, 2021). Therefore, it can be determined that HEIs require data governance roles such as data

trustees who are accountable for identifying possibilities of institutional (i.e., data use) data to augment overall mission of the institutions.

The other challenges to data governance practices at HEIs were related to the absence of clear policy. For instance, the Open science policy at HEIs focuses on making "the data machine-readable" by applying FAIR principles. Although the policy outlines the principles of good data management, data quality attributes such as relevance, accuracy, timeliness, and completeness were missing in the policy document. Therefore, the quality of research data (metadata) may suffer.

Majority of challenges in data governance are related to data value and data quality. The findings of study revealed that establishing business use of data at HEIs constrained by lack of expertise, failures in the systems, and lack of data quality, sensitivity of data (e.g. research huma being), restrictions of regulations (e.g., GDPR restrictions) pertaining data protection, and research ethics, possibilities of data leak and security threats.

The findings of this study indicated that the HEIs were struggling with maintaining data quality. As emphasized by DAMA (2017), the essence of data governance is to maintain data quality of organizations. Majority of the participants believed that the absence of data quality tools and plans, non-standard data format, subjectivity of data, human errors, complexity systems (e.g., student data system), and lack of accuracy and absence of internal data quality audits were the major inhibitors of data governance practices from the aspect of data quality decision domain. As agreed by DAMA (2017), poor data quality arises due to issues such as lack of planning, data silos, inadequate documentation, incompatible system development, absence of standards and policies, and lack of understanding the existing data assets. In addition, librarians at HEIs, particularly, indicated that data quality have been constrained by factors such as researchers' negligence to publish research metadata and papers, human errors, lack of competencies, absence of detailed policies and standards, poor leadership, and lack of trust in employees, and data conversion issues, lack of technology to manage digital data. For instance, the librarians at HEIs said that it is time-consuming task to track researchers works when they fail to follow the research data management processes and publish their papers. This may perhaps affect the quality of report generated by the research data management team.

The HEIs have been collecting and sharing their research metadata with other Finnish HEIs and national databases. The finding of the study indicated that occasionally the metadata of publications made by external parties exist for temporary period which in turn affects to open parallel publications. In the same way when researcher from the HEIs forgot to publish their metadata through the HEIs library systems, other institutions were unable to access the metadata. In line with this, IPR issues also constrained to access the publications metadata if the publications are licensed due to the concern of research funding organizations. As result, it impacts the fundamental goals of open science agenda such as openness or open data sharing and reusing. The report conducted by National Open Science Coordination, Federation of Finnish Learned Societies (2021) supported the findings of this study in stating that lack of technology or absence of infrastructure, lack of necessary skills to manage research data assets, influence of legislation and laws, and contractual restrictions, and research ethics issue are the major inhibitors for opening research data.

As mentioned above (**RQ1**), the results of this study indicated that the GDPR has impact in all data governance decision domains. As results, pertinent principles, guidelines, policies, and guidelines were designed in accordance with the GDPR. It restricts the way all academic community handle personal data. Experts have been conducting Transfer Impact Analysis (TIA) when they use cloud services under different countries jurisdiction. For majority of participants in this study the GDPR was too strict which affect to obtain the potential value from their data assets.

As mentioned above, The ICT department at HEIs are responsible for planning a data storage solution. The finding of the study revealed that the existing data storage solutions could become graveyard for old datasets due to inadequacy of data archiving policy to preserve such data assets. As highlighted by Wilkinson et al. (2016), the goal of good data management and data stewardship should not only be limited to collection and annotation of data rather should include incorporate the idea of overseeing valuable data assets for long-term. Alongside, researchers, staff and student could only gain data access rights as long as they are affiliated with the HEIs. In general, the current data storage solutions had an impact on both the research datasets and researchers.

The finding of the study also showed challenges related to data governance practices pertaining to continuous education activities. The hiddenness of data in learning management system and students' data systems, absence of technology to understand and analyse data, lack of expertise had hindered the continuous education unit to assess competence needs in the job market, measure the performance students in specific situations (i.e., during COVID-19), and examine learning preferences students. Previous study conducted by Benfeldt et al. (2020) supported the findings of this study stating that organizations suffer to generate value from the data governance practices when there is no visibility in the systems. In addition, DAMA (2017) also pointed out that lack of understanding the present data assets could cause poor data quality. Furthermore, as suggested by Person B (administration-continuous education), defining the role of data stewards in continuous education unit could fill such kinds of gaps.

Regarding data lifecycle decision domain, the absence of data inventory plans in both HEIs may create obstacle to widely understand and monitor all the institutional data assets. Consequently, HEIs could potentially face obstacles in maintaining regulatory compliances which could in turn involve financial liabilities and reputational liabilities due to inadequacy of tracking their data assets.

The finding of the study also revealed challenges pertaining to research or project data and elements of data lifecycle issues. Due to the difference among projects (i.e., differences in terms of duration, data collection intentions and data subjects), data retention and deletion decisions and rules also varies. Therefore, the HEIs potentially may potentially face difficulties in standardizing data retention and deletion policies and plans pertaining to research data.

In general, as indicated by Khatri and Brown (2010), the five data governance decision domains namely data principles, data quality, metadata, data access, and data lifecycle are connected. The findings of this study revealed that, for instance, absence of clear definition of roles such as data stewards and data trustees, and lack of establishing policies, plans and standards pertaining to data-related activities affects data quality decision domain due to lack of guidance. This in turn affects metadata decision domain.

5.2 Conclusion

The primary goal of this study is to explore challenges and opportunities of data governance practices in Finnish Higher education Institutions (HEIs).

By considering data as institutional assets and establishing policies, principles standards, procedures, guidelines, in most of data governance decision domains, to manage their data assets both HEIs obtained several benefits from their data assets such as complying with regulations and laws pertaining to data protection, developing reports and statistics to communicate with internal management and government bodies, strengthening collaboration with other HEIs and national databases through sharing, opening and reusing research data, opening parallel publications, managing monitoring data security risks, disseminating security threats among the academic community and measuring performances.

However, the HEIs developed pertinent policies and guidelines in most of data governance decision domains, the result of the study revealed that decision domains such as data quality and data lifecycle lack clear policies and guidelines. Data quality audits and data inventory plans and guidelines were not utilized to monitoring their assets. Another major challenge was related to formalizing and assigning data governance roles and responsibilities. Even though both HEIs did not have separate data governance department and data steward's role were only defined for activities related to research data management, the finding of the study showed that experts from ICT, human resources and finance have been acting as stewards. This call for formalizing data stewards in other units at HEIs. In addition, defining new data governance roles such as data trustees can enable the HEIs to overcome issues related to data governance policies from strategic aspect.

In addition, lack of data quality audits and tools can hinder the HEIs to evaluate whether the existing data suits to the intended purposes. The finding signaled that the HEIs can overcome data quality issues by utilizing tools such as checklists. Due to absence of data inventory plan that help record all data assets of the institutions, the HEIs may fail to comply laws regulations pertaining data protection. Furthermore, the existing open science policies and data archiving plans also lack to address issues data quality attributes and long-term preservation of research data. As mentioned above most of the challenges

pertaining to data governance practice area more strategic than operational. Therefore, decision makers at HEIs are required to bridge the gaps by establishing comprehensive policies, procedure and plans that covers all the data governance decision domains. Besides, existing open science policy also need to be revised to integrate data quality dimensions with the FAIR principles. In conclusion, although data governance does not necessarily require establishing department, designing a framework that suits the existing governance practices at the HEIs is necessary.

6   FUTURE RESEARCH

This study is primarily conducted for academic purpose. Based on the findings of this study, the researcher believes that further studies can be conducted by other researchers or by concerned HEIs. The following recommendations are made based on the results of the study.

Data governance is a wide research area. Each decision areas of data governance could also be examined to obtain detailed understandings regarding their effect on data governance practices. For instance, the data quality decision domain involves several researchable areas such as data quality management, data profiling, data audits, data quality metrics, and data monitoring. In this study, issues related to data quality such as data profiling and data quality tools have not been addressed appropriately due to the absence such practices in HEIs. Also, regulations and laws such as GDPR inhibit to conduct such practices since it may involve processing of personal data. Further studies may emphasis of to examine data governance decision domains such as data quality to provide detailed insights. In line with this, the researcher suggested the following topic for further research:

(1) *Opportunities and challenges in Data quality management at Higher Education Institutions*

The finding of the study revealed that data management practices had been widely utilized in the open science and research methods, research data management process, data management plan and document management system at HEIs. Therefore, the future studies may consider examining data management maturity level at HEIS. The researcher suggested the following possible topics for future research as follows.

(2) *Data Management Maturity in Finnish Higher Education Institutions*
(3) *Implementation of Data Management In Open Science and Research Projects; The Case of Finnish Higher Education Institutions*

In this study, only seven experts from ICT, administration, and libraries areas at two HEIs have been participated to address the research objectives which hamper the researcher to generalize on the topic. Along with this, the semi-structured interview was used to collect

data from the from the participants. Therefore, further studies can fill the gap by conducting research that involves wide-range of experts and stakeholders in Higher education area.

As indicated in scope and the limitations section of this study, other experts from Data Protection Office, Finance, and Human resource were not considered in this study. Majority of participants in this study mentioned that those experts have been playing role in data-related activities decisions in higher education. Therefore, future studies may consider experts mentioned above to understand the influence regulatory environments on data governance practice, and to gain in-depth information on overall data-related activities and initiatives .

In this study, the interviewed experts have been utilizing administrative, study, library, and research data. However, the aspect of social media data and pertinent professionals assigned for this role in HEIs were not taken in account due to time limitation. Future studies may consider incorporating experts from media and communication of HEIs to acquire better insights regarding how HEIs govern their social media data.

REFERENCE

Alhassan, I., Sammon, D., & Daly, M. (2018). Data Governance Activities: A comparison between scientific and practice-oriented literature. *Journal of Enterprise Information Management, 31*(2), 300–316.

Alhojailan, M. I., & Ibrahim, M. (2012). Thematic analysis: A critical review of its process and evaluation. *West East Journal of Social Sciences*, *1*(1), 39-47.

Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2018). A systematic literature review of data governance and Cloud Data Governance. *Personal and Ubiquitous Computing*, *23*(5-6),839–859.

Ansell, C., & Torfing, J. (Eds.). (2016). *Handbook on Theories of Governance*. Edward Elgar Publishing.

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data Governance as a Collective Action Problem. *Information Systems Frontiers*, *22*(2), 299-313.

Brous, P., Janssen, M., & Vilminko-Heikkinen, R. (2016). Coordinating decision-making in data management activities: a systematic review of data governance principles. In *Electronic Government: 15th IFIP WG 8.5 International Conference, EGOV 2016, Guimarães, Portugal, September 5-8, 2016, Proceedings 15* (pp. 115-125). Springer International Publishing.

Cheong, L. K., & Chang, V. (2007). The need for data governance: a case study. *ACIS 2007 proceedings*, 100.

Dahlberg, T., Nokkala, T., Heikkilä, J., & Heikkilä, M. (2017). Data Federation with a Governance of Data Framework Artifact as the Federation Tool. *Information Modelling and Knowledge Bases XXVIII*, *292*, 31.

DAMA. (2017). *DAMA-DMBOK: Data Management Body of Knowledge* (2nd ed.). Technics Publications.

Data Protection Ombudsman Office. (2023). *Data Protection Legislation*. Tietosuojavaltuutetun Toimisto. https://tietosuoja.fi/en/legislation

DLA Piper. (2023). *Data Protection Laws of the World.* DLA Piper

Hartter, J., Ryan, S. J., MacKenzie, C. A., Parker, J. N., & Strasser, C. A. (2013). Spatially explicit data: stewardship and ethical challenges in science. *PLoS Biology*, *11*(9), e1001634.

Horsman, L. (2018, July 25). *Big Data on campus*. Higher Education Today. Retrieved from https://www.higheredtoday.org/2017/12/13/big-data-campus/

IBM. (2007). *The IBM data governance blueprint: Leveraging best practices and proven technologies*. IBM.

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K. K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R. W. W., Imming, M., Jeffery, K. G., & Kaliyaperumal, R. (2020). FAIR principles: interpretations and implementation considerations. *Data Intelligence*, *2*(1-2), 10–29.

Jim, C. K., & Chang, H. (2018). The current state of data governance in higher education. *Proceedings of the Association for Information Science and Technology*, *55*(1), 198–206.

Kassen, M. (2022). Open Data Governance in Finland: Understanding the Promise of Public-Private Partnerships. In *Open Data Governance and Its Actors: Theory and Practice* (pp. 65-96). Cham: Springer International Publishing.

Khairi, M. A. (2019). Data Governance Critical success factors on university policy document. *Indian Journal of Science and Technology*, *12*(18), 1–5.

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, *53*(1), 148-152.

KMPG International. (2018). *Data governance: Driving value in healthcare. KPMG LLP*

Koltay, T. (2016). Data governance, data literacy and the management of data quality. *IFLA journal*, *42*(4), 303-312.

Krishnaswamy, O. R., & Satyaprasad, B. G. (2010). *Business research methods*. Himalaya Publishing House.

Ladley, J. (2019). *Data governance: How to design, deploy, and sustain an effective data governance program*. Academic Press.

Linneberg, M. S., & Korsgaard, S. (2019). Coding qualitative data: A synthesis guiding the novice. *Qualitative research journal*, *19*(3), 259-270.

Madison, M. J. (2020). Data governance and the emerging university. In *Research Handbook on Intellectual Property and Technology Transfer*. Edward Elgar Publishing.

Mahanti, R. (2021). Data Governance—Metrics. In *Data Governance Success (*pp 167–196). Springer International Publishing. DOI: 10.1007/978-981-16-5086-4_7

Ministry of Education and Culture. (2016). *University Act (558/2009)*.

National Open Science Coordination, Federation of Finnish Learned Societies. (2021). Open research data and methods National policy and executive plan by the higher education and research community for 2021–2025 Policy component 1 Open access to research data. *Vastuullisen Tieteen Julkaisusarja*, *3*(7). DOI: 10.23847/isbn.9789525995480

Newman, D., & Logan, D. (2006). Governance is an essential building block for enterprise information management. *Gartner Research, Stamford, CT*, *4*.

Nielsen, O. B. (2017). A Comprehensive Review of Data Governance Literature. In *IRIS: Selected Papers of the Information Systems Research Seminar in Scandinavia* (No. 8, pp. 120-133). Scandinavian Chapter of the Association for Information Systems (AIS)-Scandinavian IRIS.

OECD. (2015). Making Open Science a Reality. In *OECD Science, Technology, and Industry Policy Papers* (Vol. 25). OECD Publishing.

OECD. (2022). *Education at a Glance 2022: OECD Indicators*. OECD Publishing.

Okoro, R. (2021). *Proposed Data Governance Framework for Small and Medium Scale Enterprises (SMES)*. Minnesota State University, Mankato.

Open Science Coordination in Finland, Federation of Finnish Learned Societies. (2020). Declaration for Open Science and Research 2020–2025. *The Committee for Public Information (TJNK) and Federation of Finnish Learned Societies (TSV)*, 2.

Paananen, J.-P. (2020). *A New Data Governance Model for the Bank of Finland*. [Master's Thesis, Metropolia University of Applied Sciences]. Theseus.

Pennanen, I. (2014). *Data governance-intelligent way of managing data*. [Master's Thesis, Turku University of Applied Sciences]. Theseus.

Plaid Consulting. (2021). *Surveying - Data Governance Policy Models*, BC Council on Admissions and Transfer.

Prasetyo, H. N. (2016). A review of data governance maturity level in higher education. *Jurnal Ilmiah Teknologi Infomasi Terapan*, *3*(1).

Privacy Technical Assistance Center. (2011). *Data Governance Checklist*. U.S. Department of Education.

Publications of the Finnish Government. (2021). *Education Policy Report of the Finnish Government*. Finnish Government.

PwC. (2019). *PwC Global and industry frameworks for data governance Global and industry frameworks for data governance Click to Launch Data Governance Knowledge Series -Topic 3*. PwC Private Limited.

Rantala, S., Swallow, B., Lähteenmäki-Uutela, A., & Paloniemi, R. (2022). Forest data governance as a reflection of forest governance: Institutional change and endurance in Finland and Canada. *Environmental Science & Policy*, *136*, 751-760.

Ruohonen, J., & Mickelsson, S. (2023). Reflections on the Data Governance Act. *Digital Society*, *2*(1), 1-9.

Saunders, M., Lewis, P., & Thornhill, A. (2016). *Research Methods for Business Students* (7th ed.). Pearson Education Limited.

Smidt, N. (2021). *Upon a Data Governance framework which uses data categorization*. [Master's Thesis, Utrecht University]. UU Theses Repository.

Soares, S. (2010). *The IBM data governance unified process: driving business value with IBM software and best practices*. MC Press, LLC.

Thomas, G. (2006). *Alpha males and data disasters: The case for data governance*. Brass Cannon Press.

Thomas, G. (2020). *The DGI Data Governance Framework*. Data Governance Institute.

Tolley, H., & Shulruf, B. (2009). From data to knowledge: The interaction between data management systems in educational institutions and the delivery of quality education. *Computers & Education*, *53*(4), 1199-1206.

Weill, P., & Ross, J. W. (2004). IT governance on one page. *Available at SSRN 664612*.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., & Gonzalez-Beltran, A. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1).

APPENDICES

Appendix 1: Interview questions

Pre-interview Briefing

- The interview will take approximately one hour.
- You have the right to withdraw from the interview at any time.
- With your permission, the interview will be recorded and transcribed for further analysis.
- The transcript of the interview will be anonymized so that you cannot be traced back.
- The interview record and transcript will be deleted after the research is completed.
- The interview questions were developed based on the five Data Governance Domains (Khatri and Brown, 2010). In addition, a data governance checklist developed by Privacy Technical Assistance Center (2011) was also utilized to develop the interview question.

Background information

- Role or title
- Years of experience
- Function or team
- Highest level of education

1. **Data principles**

1. Do you have data stewards in your organization? Where do data stewards reside? Can you describe their role?
2. How do you describe the uses of data for the organization?
3. How do you describe data as asset?
4. What are the mechanisms for communicating business uses of data on an ongoing basis?

5. How do you describe internal, external and research data as data as assets? (Do you consider inter internal, external and research data sources as data asset in your organization)

6. How do you identify opportunities for data sharing and reusing?

7. How does the regulatory environment influence the business uses of data?

8. What challenges can undermine the value of data as an asset?

2. Data quality

1. Does your institution have established policies and procedures to maintain data quality in terms of data quality attributes such as timeliness, completeness, accuracy, and relevance?

2. Does your organization have data quality tools and plans?

3. How frequently are data monitoring and cleaning performed?

4. Do you conduct data quality audits? If yes, how often do you conduct data quality audits? Do you also conduct data profiling

5. How will data quality as well as the associated program be evaluated?

6. What are the challenges regarding data quality in your department/organization?

3. Metadata- helps to organize, find, and understand data

1. Do you have a program/procedure for documenting the **semantics** of data? (Description, definition of the data)

2. How does your institution define and model metadata?

3. What is the plan to keep several types of metadata up to date?

4. How do you describe the importance of data dictionary (business glossary) for data interpretation in your organization?

4. Data access

1. Do you have policies and procedures in place that restricts users' data access? Do you consider data users roles and responsibilities when you limit data access?

2. Are these policies and procedures consistent with GDPR?

3. Do you have internal procedural control such as Non-Confidential Agreement (NDA), training and security screening in place to monitor users' data access?

4. Are there policies and procedures in place to restrict and monitor data access of authorized users (e.g., researchers) to ensure the conditions of their access to data in the system are consistent with those outlined in the data governance plan, including which data elements can be accessed, for what period of time, and under what conditions?

5. Do you have backup and recovery program to prevent potential loss of data? (What is the program for backup and recovery?)

6. What is the business value of data?

7. Do you conduct risk assessment regularly?

8. How will assessment results be integrated with the overall compliance monitoring efforts?

9. What are data access standards and procedures in your organization?

10. What is the program for periodic monitoring and audit for compliance?

11. How is security awareness and education disseminated?

5. Data lifecycle

1. How is data inventoried in your organization? Is there pertinent polices to data inventory ?

2. Does your institution have program for data definition, production, retention, and retirement for distinct types of data (internal, external, and research data)?

3. How do the compliance issues related to legislation such as GDPR affect data retention and archiving?