MSc Business Information Technology
Thesis

Assessing the FAIRness of Credit Risk Assessment: A Comparative Study of Traditional Banks and P2P Lending Platforms

Noah Koerselman

Supervisors:
Marcos Machado - Faculty of Behavioural, Management and Social sciences
Patricia Rogetzer - Faculty of Behavioural, Management and Social sciences
Luis Bonino da Silva Santos - Faculty of Electrical Engineering, Mathematics and Computer Science

July, 2025

University of Twente,
Faculty of Electrical Engineering,
Mathematics, and Computer Science,
7500 AE Enschede,
The Netherlands

UNIVERSITY OF TWENTE.

Acknowledgment

I want to thank my supervisors, Marcos Machado, Patricia Rogetzer and Luis Bonino da Silva Santos, for their helpful guidance, support and constructive feedback during this study. I have learned a lot from their expertise and insights, which have been essential to the direction and quality of this study.

I also want to express my gratitude to the representatives of the financial institutions who took the time to participate in this research and share their knowledge, making this research possible.

Abstract

As financial institutions increasingly rely on data-driven decision-making for credit risk assessment, robust data management becomes essential. This thesis examines the data management practices of traditional banks and Peer-to-Peer (P2P) lending platforms by assessing their alignment with the FAIR principles. The FAIR principles are a set of guidelines to enable data to become more findable, accessible, interoperable and reusable. A systematic evaluation framework was developed and applied using both publicly available documentation and survey responses from one traditional financial institution and one alternative finance institution.

Findings reveal that neither institution fully complies with all aspects of the FAIR principles. Both traditional banks and P2P lending platforms perform well in Findability and basic Interoperability. However, significant differences arise in Accessibility. While traditional banks benefit from regulatory frameworks, their internal practices lag in areas such as open protocols and durable metadata storage, resulting in a lower alignment with Accessibility. In contrast, P2P lending platforms show stronger internal alignment with Accessibility through the use of open protocols and durable metadata infrastructure. Reusability remains a challenge for both sectors, primarily due to incomplete systematic provenance tracking.

The study concludes that while traditional banks benefit from regulatory frameworks and established governance structures, P2P lending platforms may be ahead in some aspects of FAIR alignment, particularly in internal Accessibility. The research highlights the importance of harmonized standards and suggests targeted improvements for both sectors to enhance FAIR compliance.

Keywords: FAIR principles, data governance, credit risk assessment, banks, P2P lending platforms, financial technology.

Contents

List of Figures

List of Tables

Acronyms

AI Artificial Intelligence.

API Application Programming Interface.

ARK Archival Resource Key.

BCBS Basel Committee on Banking Supervision.

DOI Digital Object Identifier.

EBA European Banking Authority.

ECSP European Crowdfunding Service Providers.

ESAP European Single Access Point.

EU European Union.

FAIR Findable Accessible Interoperable Reusable.

FIDA Financial Data Access.

FinTech Financial Technology.

FTP File Transfer Protocol.

GDPR General Data Protection Regulation.

HTTPS Hypertext Transfer Protocol Secure.

ISO International Organization for Standardization.

JSON-LD JavaScript Object Notation for Linked Data.

LEI Legal Entity Identifier.

ML Machine Learning.

OWL Web Ontology Language.

P2P Peer-to-Peer.

PSD2 Payment Services Directive 2.

RDF Resource Description Framework.

SDMX Statistical Data and Metadata eXchange.

Chapter 1

Introduction

This chapter introduces the research topic by providing background information about credit risk assessment and an introduction to the Findable, Accessible, Interoperable and Reusable (FAIR) principles. Further, the problem statement is defined, the research objectives and questions are outlined, and the scope and structure of the research are explained.

1.1 Background & Motivation

Over the past decade, the financial industry has undergone a significant transformation driven by digital innovation [137]. Advances in financial technology (FinTech), artificial intelligence (AI) and machine learning (ML) have enabled financial institutions to automate risk assessment, enhance fraud detection, improve operational efficiency, streamline regulatory compliance and optimize trading strategies [40]. This digital shift, combined with improved regulation and standardization, has accelerated the adoption of big data in financial organizations [137].

Organizations are increasingly recognizing data as a valuable asset that can provide a strong competitive advantage, as long as it is properly governed and managed [111]. However, to unlock its full potential, data requires structured management and governance [111]. Data governance is a critical business function that includes defining data ownership, establishing data quality standards and ensuring data usage aligns with the strategic objectives [111]. As the daily amount of data generated and stored grows, so does the importance of having a proper data management framework in place [144]. The ability to extract meaningful value from data is especially important in areas that rely heavily on data-driven decision-making.

For instance, in the finance domain, credit risk assessment heavily depends on data-driven decision making. In this field, access to high-quality, well-structured data offers improvements in accuracy and efficiency of the creditworthiness evaluation of individuals and businesses [127]. Credit risk assessment is the process of evaluating the likelihood that a borrower will default on a loan. Accurately determining the creditworthiness of people, companies or financial instruments is essential for minimizing the risks associated with lending [158]. Traditionally, credit risk assessment primarily relied on historical data and the subjective judgment of experts to predict the likelihood of borrower defaults [107]. However, with the rise of big data, financial institutions now have access to vast amounts of structured and unstructured data from diverse sources [59]. This expanded data landscape presents an opportunity to enhance credit risk models by offering more detailed insights into borrower behavior and potential risk indicators [127].

While big data has the potential to improve credit risk assessment, it also introduces significant challenges in effective data management. Challenges such as fragmented data, semantic heterogeneity, data privacy, lack of standardization and compliance with regulations have an effect on the quality of the data [141]. Financial institutions must navigate these challenges in order to ensure that their data is effectively managed, structured and maintained for long-term usability. Without a standardized approach to data governance, inconsistencies in data quality could hinder the reliability and fairness of credit risk assessment [123].

A potential solution to these challenges lies in the FAIR principles, which offer a structured approach to data governance and management. The FAIR principles, introduced in 2016, can be used as a set of guidelines for optimizing data management and stewardship, ensuring that data and digital assets are more easily discoverable, accessible and reusable for both humans and machines [169]. While these principles were initially developed for scientific data, they have already been adopted to some extent within multiple sectors including pharmaceuticals, healthcare and agriculture, where they have played a crucial role in ensuring standardized data management, enhancing data sharing and improving data accessibility and reusability [162, 43, 138]. Despite these successes, the adoption of the FAIR principles in the financial sector remains, to our knowledge, largely unexplored. Given that financial institutions increasingly rely on big data for credit risk assessment, incorporating FAIR principles into their data management could help address key challenges such as data fragmentation, data standardization and regulatory compliance.

While financial institutions increasingly rely on data-driven approaches for credit risk assessment, the way they manage, structure and utilize data can differ significantly. Traditional banks, as well-established financial institutions, generally follow structured risk assessment frameworks that adhere to regulatory requirements and standardized methodologies [69]. However, they often face challenges related to outdated infrastructure, constrained resources and reliance on subjective decision-making [72]. In contrast, Peer-to-Peer (P2P) lending platforms, which operate in a decentralized and digital-native environment, leverage alternative data sources and ML models to assess borrower risk. While this may enhance flexibility and inclusivity, it also raises concerns regarding data asymmetry, transparency and regulatory alignment [156].

While P2P lending platforms and banks differ in their approach to data management and credit risk assessment, it remains uncertain how well either aligns with structured data governance frameworks such as the FAIR principles. Given the growing reliance on data-driven decision-making in finance, evaluating the FAIRness of the credit risk assessment processes within financial institutions is essential to understanding the extent to which they support robust, standardized and well-governed financial data management.

1.2 Problem Statement

The increasing reliance on data-driven decision-making has transformed credit risk assessment in both traditional banks and P2P lending platforms. This transformation has brought challenges to credit risk assessment, including data privacy concerns, regulatory complexities and technical hurdles such as data integration and standardization. To address these challenges effectively a robust data management framework is essential [127].

The FAIR principles can be used as a guideline for proper data management. The FAIR principles have improved data findability, accessibility, interoperability and reusability in industries such as healthcare, pharmaceuticals and agriculture [162, 43, 138]. While these industries have benefited from the use of the FAIR principles, financial industries have yet to assess how well their credit risk assessment processes align with these standards.

Different financial institutions might adopt different risk assessment methodologies. Therefore, a difference in alignment with the FAIR principles can also be expected between traditional banks and P2P lending platforms. This research aims to bridge this research gap by assessing the FAIRness of credit risk assessment in banks and P2P lending platforms. To conduct this assessment, we will develop a systematic evaluation framework based on the FAIR principles. By doing so, this research identifies which sector, traditional banks or P2P lending, demonstrates stronger adherence to FAIR principles, finds barriers that exist to FAIR adoption and proposes recommendations to enhance data findability, accessibility, interoperability and reusability.

1.3 Research Objectives

The aim of this research is to evaluate the FAIRness of credit risk assessment in traditional banks and P2P lending platforms by systematically assessing their alignment with the FAIR principles. A systematic evaluation framework is developed to measure FAIR compliance in credit risk assessment processes and thereby identify differences between traditional banks and P2P lending platforms in their data management practices.

To achieve this research aim, the following research objectives are defined:

1. Examine the credit risk assessment process used by banks and P2P lending platforms, highlighting their differences in regulations, data sources and data management.

2. Develop a FAIRness evaluation framework to assess the alignment of the credit risk assessment processes in banks and P2P lending platforms with the FAIR principles.

3. Evaluate and compare the FAIRness of data management practices in banks and P2P lending platforms using publicly available data.

4. Identify key challenges and barriers that hinder the adoption of FAIR principles in credit risk assessment.
5. Formulate actionable recommendations for improving the FAIRness of credit risk assessment practices in financial institutions.

1.4 Research Questions

To systematically evaluate the FAIRness of credit risk assessment in traditional banks and P2P lending platforms, this research is guided by the following main research question and a set of sub-research questions, based on the research objectives, that address specific aspects of the research.

> *Which credit risk assessment process, that of traditional banks or P2P lending platforms, demonstrates greater alignment with the FAIR principles?*

Sub-Research Questions:

SRQ1: What is the current credit risk assessment process used by traditional banks and P2P lending platforms, and how do they differ in data management practices?

SRQ2: What framework can be developed to systematically assess the FAIRness of credit risk assessment?

SRQ3: To what extent do banks and P2P lending platforms comply with FAIR principles in their credit risk assessment practices?

SRQ4: What are the main barriers preventing banks and P2P lending platforms from compliance with the FAIR principles in credit risk assessment?

SRQ5: How can the FAIRness of the credit risk assessment process in banks and P2P lending platforms be improved?

1.5 Scope

The research focus is on evaluating the FAIRness of credit risk assessment in traditional banks and P2P lending platforms, comparing their alignment with the FAIR principles by developing a systematic evaluation framework. Given this scope of the research, several delimitations are established. These can be divided into institutional scope, geographical scope, data scope, FAIR assessment scope, regulatory and legislative scope and technological scope.

- **Institutional Scope**: This research compares traditional banks and P2P lending platforms. Other types of financial institutions are not explicitly analyzed.

- **Geographical Scope**: Financial institutions are active over the entire world. However, at different locations different regulations apply. This research focuses on European banks and P2P lending platforms. Therefore, all location related aspects, such as regulations and guidelines, are considered from a European perspective.

- **Data Scope**: Since this research is conducted independently without direct access to internal company data, the study relies on publicly available data to assess the FAIRness of credit risk assessment data. However, some non-public input was obtained through a survey to complement the analysis. Proprietary and confidential credit risk assessment data from banks or P2P platforms are not included.

- **FAIR Assessment Scope**: The research does not focus on evaluating the performance of credit risk models. Instead, it assesses how well credit risk data aligns with the FAIR principles, regardless of the performance of the models used.

- **Regulatory and Legislative Scope**: While regulations and legislation influence data management, this study does not perform an in-depth legal analysis. Instead, it explores how the FAIR principles can help meet regulatory requirements, such as the General Data Protection Regulation (GDPR[1]) and the Data Governance Act (DGA[2]). However, it does not provide a legal evaluation of compliance obligations or potential legal risks.

- **Technological Scope**: The study focuses on the structural and governance aspects of FAIR data management instead of developing a new credit scoring algorithm. The study provides recommendations on how existing data governance frameworks can be improved by integrating FAIR principles, without delving into the technical implementations of ML models.

These boundaries clearly define the research scope, outlining what the research does and does not cover. This provides direction and ensures that the research stays focused, feasible and aligned with the objectives.

1.6 Methods

To answer the research questions and achieve the stated objectives, this study adopts a comparative, mixed-methods approach combining both qualitative and quantitative data collection and analyses. The research starts with a literature review exploring the current credit risk assessment processes and data management practices within traditional banks and P2P lending platforms. This is followed by the development of a FAIRness evaluation framework. The developed framework is used to answer the research question. With the developed framework, the FAIR alignment of the credit risk assessment process within banks and P2P lending platforms is scored

[1]GDPR: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng
[2]DGA: https://eur-lex.europa.eu/eli/reg/2022/868/oj/eng

using a set of defined evaluation criteria. To represent the data processes of the institutions, publicly available literature and documentation are used. The structured scoring enables the identification of potential improvements in FAIR alignment and supports answering the main research question: which credit risk assessment process, that of traditional banks or P2P lending platforms, demonstrates greater alignment with the FAIR principles?

1.7 Outline

This thesis consists of eight main chapters. Together, these chapters contribute to a systematic evaluation of the FAIRness of credit risk assessment in traditional banks and P2P lending platforms. These chapters are as follows:

- **Chapter 1 - Introduction**: Introduces the research topic, defines the problem statement, outlines objectives and questions, and explains the scope of the study.
- **Chapter 2 - Theoretical Background**: Presents the FAIR principles and their sub-principles as the conceptual foundation of the research.
- **Chapter 3 - Literature Review**: Reviews relevant literature on credit risk assessment practices and data governance in traditional banks and P2P lending platforms.
- **Chapter 4 - Methodology**: Describes the research design, data sources, and evaluation approach.
- **Chapter 5 - The FAIR Alignment Evaluation Framework**: Details the framework used to assess FAIR alignment.
- **Chapter 6 - Results and Analysis**: Presents the findings from applying the evaluation framework.
- **Chapter 7 - Discussion**: Interprets the results and reflects on their implications.
- **Chapter 8 - Conclusion**: Summarizes key findings and offers recommendations for practice and future research.

This structured approach ensures a logical progression of ideas, guiding the reader from theoretical foundations to practical findings.

Chapter 2

Theoretical Background

This chapter provides a theoretical background for this research. It explores the FAIR principles to establish a basis for better understanding of the research problem. This is done by introducing the principles and breaking them down to their sub-principles.

2.1 Introduction to the FAIR principles

The FAIR guiding principles as originally proposed by Wilkinson et al. [169] stand for Findable, Accessible, Interoperable and Reusable. These principles are divided in fifteen sub-principles, for more details see Table 2.1. The goal of pursuing the FAIR principles is to enhance the quality of data by ensuring it is easily discoverable and reusable [169]. To help achieve these goals, the principles serve as a guideline for optimizing data management and ensuring that data is structured, maintained and shared in a way that enhances its usability for both humans and machines.

The FAIR principles, published in 2016, were originally introduced to improve scientific data management and stewardship. Since then a growing number of organizations across various sectors have actively tried to make technical implementation decisions that adhere to the FAIR principles [125]. The FAIR principles are aspirational, offering organizations the flexibility to adopt, adapt, or prioritize specific aspects of them based on their needs, allowing for a tailored approach to improving data management practices [161].

To better understand the significance of FAIR in data management, section 2.2 provides a detailed breakdown of each principle, explaining its role in structuring and optimizing data.

2.2 Breakdown of the FAIR principles

As mentioned earlier the FAIR principles consist of four core elements, Findable, Accessible, Interoperable and Reusable, each with a set of sub-principles that guide best practices in data management. This section describes each of these core elements along with their respective sub-principles in detail to get a better understanding of their role in enhancing data management practices.

TABLE 2.1: The FAIR Guiding Principles by Wilkinson et al. [169]

FAIR Guiding Principles
Findable Data should be easy to locate for both humans and machines. This requires proper documentation and the use of persistent identifiers. **F1**: (meta)data are assigned a globally unique and persistent identifier. **F2**: data are described with rich metadata (defined by R1 below). **F3**: metadata clearly and explicitly include the identifier of the data it describes. **F4**: (meta)data are registered or indexed in a searchable resource.
Accessible Once found, data must be retrievable using standardized, secure protocols. **A1**: (meta)data are retrievable by their identifier using a standardized communications protocol. **A1.1**: the protocol is open, free, and universally implementable. **A1.2**: the protocol allows for an authentication and authorization procedure, where necessary. **A2**: metadata are accessible, even when the data are no longer available.
Interoperable Data should integrate with other data and systems using standard formats and languages. **I1**: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. **I2**: (meta)data use vocabularies that follow FAIR principles. **I3**: (meta)data include qualified references to other (meta)data.
Reusable Data should be well-described and curated so it can be replicated and combined in different contexts. **R1**: meta(data) are richly described with a plurality of accurate and relevant attributes. **R1.1**: (meta)data are released with a clear and accessible data usage license. **R1.2**: (meta)data are associated with detailed provenance. **R1.3**: (meta)data meet domain-relevant community standards.

2.2.1 Findability

For data to be useful it must in the first place be easily discoverable [13]. Data should be easy to find by both humans and machines [13]. The Findability principle emphasizes the use of unique identifiers, rich metadata and indexing in searchable resources.

F1: (meta)data are assigned a globally unique and persistent identifier. Principle F1 is arguably the most important principle, since globally unique and persistent identifiers are essential elements found in all of the other FAIR principles [100]. An identifier is globally unique, sometimes also called universally unique, when it identifies an entity without the dependence on a central registration authority. This means that the identifier unambiguously refers to one entity in the world, without being limited to a specific system or context. An identifier that is only locally unique is insufficient, as it may lead to conflicts or ambiguities when used outside its original scope [100]. In addition to being globally unique, the identifier should also be persistent. The persistency of the identifier refers to the ability of the identifier to remain stable and valid over time. Persistence requires clear polices for managing changes in the environment [65]. Examples of globally unique and persistent identifiers are Digital Object Identifiers (DOIs) and Archival Resource Keys (ARKs) [13].

F2: data are described with rich metadata (defined by R1 below). Metadata is a set of data that describes and gives information about other data. The use of metadata is important for describing, categorizing and organizing data, which allows for the data to be easily found again [89]. The metadata should be generous and extensive, including descriptive information about the context, quality and condition, or characteristics of the data [13]. The more detailed the metadata, the easier it is to find and understand the data.

F3: metadata clearly and explicitly include the identifier of the data it describes. This principle states that the metadata is not isolated but remains connected to the dataset. The metadata and the dataset it describes are usually separate files. This means that the association between the metadata file and the dataset should be made explicit by mentioning a dataset's globally unique and persistent identifier in the metadata [13]. Including the identifier in the metadata ensures that humans and machines can reliably connect metadata records to the correct dataset. Without this connection the metadata can become detached from the dataset it describes, making retrieval difficult.

F4: (meta)data are registered or indexed in a searchable resource. This principle states that data and its metadata should be registered in a system that allows it to be searched and discovered by humans and machines. This searchable resource provides the infrastructure to search for the data [167]. Indexing ensures that the data does not exist in isolation and that it can be found in a searchable repository by the use of a query. Without indexing, even well documented data can be underutilized.

2.2.2 Accessibility

Once the required data is found, ensuring that the data is accessible is the next step towards FAIR data. The user needs to know how the data can be accessed. FAIR does not mean that the data must be openly available to everyone. Instead, the idea of FAIR data is that the data should be as open as possible and as closed as necessary, thus accessible under clear conditions [117].

A1: (meta)data are retrievable by their identifier using a standardized communications protocol.
For data to be accessible, it must be retrievable via a standardized communication protocol. Specialized or proprietary tools should not be required for retrieving the data [13]. A communications protocol is a set of formal rules describing how data is transmitted or exchanged between different devices. Standardized means that the communication protocol follows a globally used and recognized set of rules [164]. Examples of such standardized data retrieval protocols include Hypertext Transfer Protocol Secure (HTTPS) and File Transfer Protocol (FTP) [119]. Such a protocol enables both manual retrieval by users as well as automated access by machine. In cases of highly sensitive data fully mechanized access is not always possible. In such scenarios, it is perfectly FAIR to use non mechanized access protocols, such as manual requests to a data custodian, as long as the protocol is clearly defined in the metadata [13].

A1.1: the protocol is open, free, and universally implementable.
To ensure broad accessibility, the data retrieval protocol should be based on open standards rather than proprietary or restricted systems. Furthermore, the protocol should be free of charge [13]. Therefore, there should be no dependency on specialized tools or paid services. Anyone should be able to at least access the metadata, regardless of the user's platform, location or institutional affiliation [13].

A1.2: the protocol allows for an authentication and authorization procedure, where necessary.
This principle clearly states that the data does not have to be openly available to everyone. Instead, clear conditions must be provided under which the data is accessible. If a dataset contains sensitive or proprietary information the dataset can still be FAIR as long as the accessibility conditions are explicitly defined [13]. In such cases, proper authentication and authorization mechanisms, such as OAuth, API keys, and institutional logins, should be in place to regulate the access while maintaining security [96]. Additionally, accessibility requirements should be machine-readable, since humans as well as machines should be able to request the data. This enables automated systems to interpret access restrictions and either execute the necessary authentication procedures or alert the user about specific requirements for data retrieval [13].

A2: metadata are accessible, even when the data are no longer available. Data may not always be permanently stored. Datasets may become outdated, deleted due to storage limitations or too high maintenance costs or become unavailable due to policy changes. When a dataset is no longer accessible, its metadata should still persist, preventing the search for data that may no longer exist. Datasets that have been referenced by others should retain their metadata so that future users can understand their context, content and provenance [100].

2.2.3 Interoperability

Data interoperability is essential for ensuring that the data can be used by different systems [154]. Data that is stored in inconsistent formats can create data silos, which prevents effective collaboration [154].

I1: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
For data to be interoperable, the data must be both human readable and machine readable, adhering to widely accepted standards. This ensures that different systems can process the data without the need for specialized translators or mappings [13]. This principle establishes the need for a standardized knowledge representation language to understand data across different domains. This language should accurately describe and define objects, so that machines can distinguish between different meanings of the same term [100]. These languages should have a few properties. First these languages should have a formal specification, meaning its syntax and grammar are precisely defined. Secondly, its specifications should be shared and accessible, allowing others to read the specifications and learn the language. Thirdly, the language must be designed for use in more than one scenario rather than being restricted to a single context [13]. Examples include Resource Description Framework (RDF), JavaScript Object Notation for Linked Data (JSON-LD) and Web Ontology Language (OWL) [13].

I2: (meta)data use vocabularies that follow FAIR principles.
For data to be interoperable, it must use consistent and standardized vocabularies so that different users, humans as well as machines, can understand and interpret the data in the same way. The vocabulary provides terms or concepts to represent the data [13]. Without standardized terms, the same concept might be labeled differently across datasets, making data integration difficult. The vocabulary itself should also be findable, accessible, interoperable and reusable [100]. An example of such a FAIR-compliant vocabulary is the DDI-RDF Discovery Vocabulary provided by the DDI Alliance, which offers standardized, machine-readable terms [12].

I3: (meta)data include qualified references to other (meta)data.
For data to be interoperable, datasets should link to related datasets so that humans and machines can easily find relevant, connected information. Following this principle means that the reference should explicitly describe the nature of the relationship. A qualified reference should specify whether one dataset builds on another dataset,

whether additional datasets are needed to complete the data, or whether additional information is stored in a different dataset [13]. This ensures correct interpretation of the connection.

2.2.4 Reusability

For data to be effectively reused, it must be well described [13]. Reusability is the main goal of the FAIR principles, aimed at maximizing the value of data by ensuring that it can contribute to future research efforts [152].

R1: meta(data) are richly described with a plurality of accurate and relevant attributes.
This principle states that the data should come with detailed information that helps others understand and use it properly. This principle looks similar to F2, however, there is a difference. While F2 focuses on making the data searchable by providing specific attributes, R1 ensures that the data includes detailed descriptions that enable the users to interpret and use the data correctly [13]. Following this principle means that the metadata should not only describe the content of the data but also the context in which it was generated. This includes, among others the scope of the data, versioning and details about the collection, conditions and limitations of the data [13]. Further, the term plurality emphasizes that metadata should be as rich and inclusive as possible, even including information that may seem irrelevant [13].

R1.1: (meta)data are released with a clear and accessible data usage license.
To ensure reusability, the data should be explicitly licensed using Creative Commons, Open Data Commons or similar frameworks [116]. The usage rights of the data should be described clearly. Unclear or ambiguous licensing can create significant barriers to data reuse. A well-defined license ensures that both humans and machines can interpret the conditions under which the data can be used, modified and shared [13].

R1.2: (meta)data are associated with detailed provenance.
Provenance refers to the history and origin of the data. Provenance information should include information about the creation of the data, who contributed to it, how it has been processed and what transformations the data has undergone [91]. Provenance details enable users to assess the accuracy and timeliness of the data [63].

R1.3: (meta)data meet domain-relevant community standards.
In case community or domain-specific standards exist, they should be followed. These standards ensure consistency within their domain. Examples are metadata templates, controlled vocabularies, minimal information checklists and ontologies [126]. In some cases, deviations from these standards are possible, provided there is a good reason and this is addressed in the metadata [13]. The FAIR principles focus on data usability and structure, rather than assessing data quality.

2.3 Summary

This chapter introduced the FAIR principles, which serve as guidelines for improving data management and stewardship. Each sub-principle was explained in detail to provide a full understanding of the framework. The principles emphasize the importance of making data easily discoverable, securely accessible, interoperable across systems and reusable over time. Understanding these principles is essential to later understand the criteria used to evaluate credit risk assessment processes within financial institutions. These principles are further contextualized in Chapter 5, where they form the basis of the FAIRness evaluation framework.

Chapter 3

Literature Review

This chapter includes a literature review of credit risk assessment of banks and P2P lending platforms. By exploring regulations, data sources and data management approaches, this review highlights how these institutions handle credit related information. Examining and comparing these data management practices lays the foundation for evaluating the FAIRness of credit risk assessment in these financial institutions.

Credit risk assessment involves evaluating the probability of a borrower failing to meet their loan obligations. How suitable a borrower is to receive financial credit is called creditworthiness. Effectively determining the creditworthiness of a borrower is vital for financial institutions, as it helps mitigate risks associated with lending and minimizes financial losses [158].

The process of credit risk management generally consists of four key steps: identification, measurement, management and control [106]. Each financial institution implements these steps in different ways depending on its business model, regulatory requirements and risk tolerance. This results in different credit risk assessment processes for different financial institutions. While multiple types of financial institutions engage in lending, this research focuses specifically on banks and P2P lending platforms, as they represent two distinct approaches to credit risk assessment.

3.1 Credit Risk Assessment

The following sections explore the credit risk assessment in banks and P2P lending platforms for personal loans, highlighting their key differences in regulations, data sources and data management practices. This comparison establishes the necessary foundation for understanding the data management practices used by these financial institutions, so that later these practices can be examined in more depth to assess the alignment with the FAIR principles.

3.1.1 Banks

Traditional banks are the most common form of banking and refer to highly regulated financial institutions that provide a wide range of financial services, including private banking, checking and savings accounts and loan services [17]. Historically, these banks have a physical location where customers can access financial services, but with the rise of FinTech many now offer extensive mobile and online banking services [121]. One of the cornerstones in banking operations is credit risk assessment, ensuring financial stability and profitability [36]. The credit risk assessment process in banks is highly standardized and subject to strict regulations to ensure the safety and stability of banks.

Regulations

Banks must comply with international regulatory frameworks, such as the Basel Accords[1], guidelines set by the European Banking Authority (EBA)[2], European Union directives[3] and European Union regulations[4]. These regulatory frameworks aim to ensure the stability of the financial system.

The Basel Accords (Basel I, Basel II, Basel III) are a set of sequential international regulation agreements developed by the Basel Committee on Banking Supervision (BCBS) to ensure that banks have enough capital to meet their obligations and handle unexpected losses [79]. Basel I defined guidelines for the minimum ratio of capital to risk-weighted assets, which was set at 8%. Basel II refined and added on Basel I by introducing standardized measurements for credit risk, market risk and operational risk. Basel III revises and strengthens the three pillars of Basel II and extends them in several areas [52]. Compliance with these accords ensures that banks are resilient to financial shocks and can maintain stability in adverse economic conditions.

In addition to international banking regulations the EBA, established in 2011, harmonizes regulatory practices across European Union (EU) member states. The EBA is a regulatory authority that provides guidelines, recommendations and technical standards to ensure uniform banking regulation across the EU [66]. It plays a key role in promoting effective and consistent prudential regulation and supervision across European banks. Additionally, it provides regular stress tests and oversees the implementation of the Basel Accords within the EU [99]. One of the main guidelines for credit risk assessment is the EBA Guidelines on Loan Origination and Monitoring (EBA/GL/2020/06)[5] [81]. These were introduced in June 2021 and set common standards for how banks should grant and monitor loans. The guidelines cover requirements for borrowers' creditworthiness assessment, data requirements, internal governance, loan origination procedures and credit monitoring. They also bring together the EBA's prudential and consumer protection objectives [81].

Regulatory oversight within the EU extends further through EU regulations. EU regulations and directives are two primary forms of legislation used by the EU to

[1] Basel Framework: https://www.bis.org/basel_framework/

[2] EBA Guidelines on loan origination and monitoring: https://www.eba.europa.eu/activities/single-rulebook/regulatory-activities/credit-risk/guidelines-loan-origination-and-monitoring

[3] European Union directives: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=legissum:l14527

[4] European Union regulations: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=legissum:l14522

[5] EBA/GL/2020/06: https://www.eba.europa.eu/activities/single-rulebook/regulatory-activities/credit-risk/guidelines-loan-origination-and-monitoring

create legal frameworks across member states. EU regulations are binding legislative acts in all member states, so they do not require national governments to pass additional legislation [122]. An example of an EU regulation is the General Data Protection Regulation (GDPR)[6], which enforces protection and privacy of personal data across the EU. On the other hand, EU directives set legal objectives that member states must transpose into national law in order to achieve them, allowing flexibility in how they are implemented [122].

Beyond these overarching regulations, national supervisory authorities impose additional risk assessment requirements tailored to local economic conditions. These regulations collectively ensure that banks conduct thorough and responsible credit risk assessments, mitigating financial risk while maintaining consumer protection.

Data Sources

Banks apply various models to assess the creditworthiness of borrowers. These models can broadly be divided into three categories, which are human expert-based models, statistical models, and ML models [64]. The accuracy and reliability of these models depends heavily on the quality and variety of data sources used by banks. To form a complete view of the borrower's financial profile, banks use a combination of internal and external data sources.

The minimal data that should be available for the creditworthiness assessment is the purpose of the loan, employment, source of repayment capacity, household composition, financial commitments and expenses for their servicing, regular expenses, collateral (for secured lending) and other risk mitigants [81]. This data is typically internally stored and is collected directly through loan applications, onboarding processes and the transactional records of the bank.

Complementing internal data, banks also use external data sources. The most notable are credit bureaus, which collect and store borrowers' credit data and provide this data in clear credit reports and credit scores. These credit reports provide credit issuers with the information they need to accurately determine credit approval for borrowers [10]. To ensure the quality of the data, credit bureaus in Europe are generally subject to regular audits and follow strict update routines, often refreshing data on a daily or weekly basis [140]. Additionally, retention periods for negative credit events such as defaults are standardized to some extent. For instance, default data is typically retained for a minimum of three years in many European countries [140]. However, there is a significant difference between countries in retention periods, which challenges cross-border consistency and reusability [140].

In recent years, open banking regulations, such as Payment Services Directive 2 (PSD2)[7] in the European Union, has expanded the potential range of external data sources. Banks can through the use of open application programming interfaces (APIs) get access to real-time financial data from accounts held at other institutions [110]. This is particularly useful when the borrower is new to the bank or lacks a traditional credit footprint.

If needed, banks may also verify information through third-party checks, such as

[6]GDPR: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

[7]PSD2: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32015L2366

contacting a borrower's employer or consulting public authorities in case there are concerns about the accuracy or reliability of the provided information [81].

Despite their access to a wide range of financial data, many traditional banks still face challenges related to data fragmentation and siloed storage systems [38]. Legacy IT infrastructures often result in data being stored across different departments or platforms in incompatible formats. These structural issues complicate efforts to streamline the credit assessment process. Moreover, such fragmented data environments pose challenges to aligning with data management frameworks such as the FAIR principles.

Data Management Practices

Traditional banks manage vast amounts of financial data, this requires effective data management for accurate credit risk assessment and regulatory compliance [108]. Banks recognize the need for good data governance. To help with this the Basel Committee published the "Principles for effective risk data aggregation and risk reporting" (BCBS 239) [53]. BCBS 239 outlines fourteen principles aimed at enhancing banks' ability to aggregate and report risk data. In practice, this has driven banks to establish rigorous data governance and thorough data management frameworks [134].

Despite the good intentions of BCBS 239, most banks do not yet fully comply with the principles [134]. Legacy systems are still widely used, and data silos continue to persist, creating significant barriers to implementation [38]. These structural challenges hinder the development of enterprise-wide data standards and complicate efforts to achieve the comprehensive risk data aggregation and reporting envisioned by the principles. To address these challenges, banks are investing in data warehouses and data lakes [93]. These are central repositories with great scalability that can be used to store structured and unstructured data.

To enhance interoperability, banks often adopt international standards. The International Organization for Standardization (ISO) provides a common framework for the best way of doing something [19]. Examples are ISO 8000, which supports data quality management, ISO 27001, which ensures secure information handling and ISO 20022, which includes standards for financial messaging [3, 20, 21]. Additionally, identifiers like the Legal Entity Identifier (LEI) promote semantic consistency and cross-system integration by uniquely identifying entities across datasets [53].

Regulations and industry guidelines have forced banks to improve their data management practices. Thereby, indirectly aligning their data management with the FAIR principles.

3.1.2 P2P Lending Platforms

P2P lending platforms provide the possibility to obtain loans from individuals through online services that match lenders with borrowers [175]. P2P lending platforms typically work as follows: the process starts with the borrower requesting a loan through the online interface. The borrower is assigned a risk level by the platform based on the borrower's profile, this effects the interest rate. The borrower then receives loan offers from lenders based on the risk level. These lenders have deposited a sum of money into their accounts to fund their loans. The borrower can accept one of these offers or multiple, this way the loan is financed collectively [31]. The transfer of loans and interests is handled by the platform [175]. This way of borrowing is mainly popular among borrowers with low- and mid-level credit ratings, a group that has difficulty getting bank loans since the global financial crisis [174]. P2P lending platforms still follow a credit risk assessment process that shares similarities with banks. However, significant differences exist in regulations, data sources and data management practices.

Regulations

While P2P lending platforms share some regulatory obligations with traditional banks, their regulatory frameworks differ significantly. Their unique structure challenges existing credit laws, often placing them within a legal vacuum where clear regulatory guidance is lacking [86]. Furthermore, regulatory approaches remain inconsistent across Europe, with individual member states maintaining their own national rules [86].

A major regulatory challenge arises from the fact that P2P lending platforms are typically not classified as creditors. Instead, they function as intermediaries that match borrowers with investors or lenders [86]. As a result, P2P lending platforms fall outside the scope of key European credit directives, such as the Consumer Credit Directive and the Mortgage Credit Directive [86]. These directives apply to institutions that provide credit directly to consumers, which excludes platforms that do not themselves lend or take on credit risk. This legal distinction contributes to a fragmented regulatory environment, as P2P lending platforms often do not have to meet the obligations that apply to traditional financial institutions.

To create some consistency across Europe, the EU Regulation 2020/1503 on European Crowdfunding Service Providers (ECSP) was created [24]. In the EU, this is the primary regulation governing P2P lending [104]. This regulation includes lending-based crowdfunding, which includes facilitating of granting of loans as well as including services such as presenting offers to clients and assessing the credit risk [24]. Key requirements of this regulation for P2P lending platforms operating in the EU are that they need to be licensed, have robust internal processes and methodologies in place, adhere to strict consumer protection and transparency rules and submit to supervision by competent authorities [24]. Although the ECSP regulation marks an important step toward regulatory harmonization, its focus remains primarily on crowdfunding for business purposes. Therefore, it does not fully cover consumer focused P2P lending, and gaps still remain. Studying these regulatory gaps helps evaluate how they affect data management practices and the alignment with the FAIR principles.

In practice, many P2P lending platforms are supervised by national financial supervisory authorities in the countries where they operate. These authorities are responsible for licensing and overseeing a wide range of financial institutions [104]. For example, Bondora[8], a platform examined in this study, is licensed as a Credit

[8]Bondora: https://www.bondora.nl/en/

Provider by the Estonian Financial Supervision Authority, and is also subject to oversight by regulators in Finland [18]. While these national financial supervisory authorities form a baseline of financial accountability, the enforcement of supervisory practices varies by country, contributing to inconsistencies in regulations across Europe.

Although P2P lending platforms often fall outside the scope of credit laws, they are still subject to general data protection and privacy laws, most notably the GDPR [86]. Since platforms collect, process and store personal data from borrowers and lenders, they must ensure that they do this in compliance with GDPR requirements. Failure to meet these obligations can result in significant penalties [24].

This regulatory environment for P2P lending platforms is less mature and consistent than that of banks. This regulatory environment influences P2P lending platforms' data management and stewardship.

Data Sources

P2P lending platforms use a variety of data sources to assess credit risk, some of which overlap with those of traditional lenders, but others are new. Like banks, P2P lending platforms use conventional credit data such as credit bureau records and their credit scores, employment, income, credit history and loan related information [148]. In addition, P2P lending platforms assign their own credit grade to the loans, which is determined by the platforms' algorithm based on the loan and borrower characteristics [148]. This platform-assigned credit grade is a key determinant of loan performance. All these determinants are typically disclosed so that investors can make informed decisions.

Next to the traditional credit risk metrics, P2P lending platforms also incorporate alternative data, which is mainly used to evaluate borrowers who lack extensive credit histories [103]. Alternative data includes data beyond the traditional financial data, such as non-financial payment streams including energy utilities and telecommunication services, mobile phone usage and social media footprints [41, 114]. The use of such alternative data, coupled with ML models, can paint a more complete picture of an individual's creditworthiness and ability to meet the individual's obligations.

Data from borrowers is typically collected through a self reporting form, where the borrower is required to provide personal, financial and employment related information [165]. To be able to verify this information the borrower is obliged to upload additional supporting documents such as an identity card, payslips and credit reports. Platforms use identity verification tools and fraud detection systems to validate user-submitted data and reduce the risk of misrepresentation [173].

Uploaded data from the borrower is processed through the platform's own risk assessment models. P2P platforms often use automated, algorithm-driven systems that weigh a variety of variables for the risk assessment [61]. The advantage of these models, which use many variables, is that they can provide access to credit at lower cost for creditworthy individuals who have a thin credit history [103]. However, the reliance on alternative data also raises concerns about disparate treatment and privacy [103]. Regarding the loans, P2P lending platforms typically maintain detailed loan performance datasets. Some large platforms have made these loan datasets publicly available in an anonymized form for investors and researchers [129]. This openness is relatively unique in finance and contributes to the transparency of the platforms.

Data Management Practices

Data management practices among different P2P lending platforms vary significantly depending on their business model, technological maturity and regulatory environment [54]. Unlike traditional banks, which have to follow strict regulations that influence their data governance significantly, P2P lending platforms typically operate with greater flexibility and fewer standardized frameworks [157]. While this flexibility supports rapid innovation and agility, it may also lead to inconsistencies in data practices.

These platforms are typically digital-native by design, meaning that data collection, processing and storage are managed entirely through online systems. To support these operations, most P2P lending platforms rely on cloud computing infrastructure. Cloud services not only enable efficient, centralized storage but also allow platforms to perform real-time analysis of borrower profiles to assess creditworthiness [112]. In addition, cloud computing provides the scalability needed to accommodate rapid growth in user bases without requiring significant investment in physical infrastructure [112].

To further enhance interoperability and operational efficiency, P2P lending platforms commonly adopt APIs as a core architectural component. Oladinni and Adewale's [131] work shows that APIs enable seamless integration of third-party services such as credit bureaus with internal systems such as risk assessment engines. This interoperability enables smooth data flow across components, streamlining workflows and improving coordination among stakeholders. For example, a lending platform can automatically retrieve a borrower's credit report through an API and use the data directly in its risk assessment model, eliminating manual steps and thus accelerating loan processing. The integration of APIs with data-driven modeling has significantly streamlined processes, reducing processing times and enhancing overall platform responsiveness. In addition, APIs support the incorporation of advanced modeling techniques, such as ML for credit scoring and risk analysis, allowing for more personalized lending experiences. By enabling modular and scalable system design, APIs play a critical role in supporting the rapid growth and flexibility of FinTech lending platforms, particularly when compared to more rigid infrastructures found in traditional banking [131].

P2P lending platforms rely on advanced and flexible data infrastructures that allow them to scale rapidly and cost-effectively. However, the absence of standardized data management practices continues to distinguish them from traditional banks, where such standards help ensure data quality, regulatory compliance and long-term operational resilience.

3.2 An overview of recent studies related to the FAIR principles

To the best of our knowledge, there is no research that explicitly applies the FAIR principles in the financial domain. However, these principles have already been successfully implemented in various other data intensive sectors to enhance data management and optimize workflows. Examples of such sectors include agriculture [160, 162], astronomy [135], pharmacy [43, 44, 45] and healthcare [46, 47, 92, 97, 98, 139, 149, 151]. Lessons from these fields provide valuable insights into the potential benefits and challenges of FAIR implementation in financial services.

3.2.1 Benefits of the FAIR Principles

The FAIR principles offer numerous advantages by promoting best practices in data management. They emphasize standardization, comprehensive documentation and improved accessibility, thereby fostering the creation of sustainable, efficient and collaborative data ecosystems [46, 124]. Moreover, the emphasis on thorough documentation and accessibility facilitates reproducibility, which is essential for verifying results and building trust in the outcomes [142]. For instance, applying the findability principles makes that data can be made more easily found and reused by using standardized metadata, unique identifiers and common data models, reducing the time spent on cleaning and transforming data [46, 124].

Accessibility in FAIR does not imply open data but ensures data is retrievable under well-defined, secure conditions. FAIR compliant access protocols allow for secure access to data, safeguarding data privacy and maintaining robust security standards [46]. This aspect of the FAIR principles ensures that organizations comply with legal and regulatory standards, such as the GDPR, while still allowing data to be available for analysis.

Interoperability plays a vital role in facilitating integration across systems [169]. The adoption of standardized data models facilitates more effective and efficient research and decision-making processes by enabling data from many sources to be seamlessly integrated, compared and analyzed in accordance with the interoperability principle [46, 97]. In credit risk assessment, interoperability could enhance data integration from credit bureaus, financial institutions and government databases.

Reusability is achieved by clear usage policies and detailed metadata documentation. This enhances datasets' long-term value by enabling them to be reused for various research applications [46]. Reusing data means that more data is available since the data is not thrown away, therefore accelerating research discovery [142]. Another benefit of reusable data is that it avoids spending money on collecting data that already exists [142].

3.2.2 Challenges to Implementation

Despite the substantial benefits of the FAIR principles, their implementation comes with notable challenges. These challenges can be found in the financial, technical, legal, organizational and cultural dimensions of implementation [43].

Regulations and legislations can complicate the process of data sharing by requiring additional considerations to ensure compliance [55]. Such regulations often make organizations more reluctant to share data due to concerns about compliance and potential sanctions [43, 49, 139, 162]. Furthermore, the lack of clear legal frameworks governing data ownership and accessibility rights lead to uncertainty about who has the authority to share or reuse data [162].

Technically, many organizations lack the infrastructure and resources required to implement FAIR-compliant systems. Data is often stored in formats optimized for internal use, which may not align with the FAIR standards [74]. Additionally, data is often fragmented across departments or systems, which leads to redundant storage, inconsistencies and inefficiencies in data sharing processes [98]. These limitations are rooted in the fact that often existing data architectures were not designed with FAIR compliance in mind [74]. Transforming them into interoperable and machine-readable formats to meet FAIR standards requires significant effort and investment in both tooling and methodological support [43, 74].

Many organizations do not prioritize FAIRification. Many organizations view these FAIRification efforts, such as the implementation of metadata schemas and data models, as specialized tasks rather than a fundamental component of their data management processes [55, 74, 139]. Without recognition, organizations are less likely to invest significant time and effort required to adopt FAIR principles, particularly when the value of these efforts is not immediately apparent [98].

Financial constraints further complicate the implementation of the FAIR principles. Implementing FAIR principles requires investments in infrastructure, personnel, training and ongoing data stewardship [43, 74, 149].

Ethical and cultural barriers also play a role. FAIR encourages data availability and reuse, yet this must be balanced against concerns around privacy, confidentiality and ethical use [49]. Financial data is inherently sensitive, and institutions may fear reputational or competitive risks associated with data sharing [55].

3.2.3 Addressing the Challenges

To address these challenges, several studies have proposed FAIRification workflows, outlining step-by-step processes supporting the full adoption of the FAIR principles [75, 101, 151, 168]. To ensure the successful implementation of the FAIR principles in data management, it also helps to adhere to established best practices. The study of Alvarez-Romero et al. [47] outlines a wide collection of best practices for the adoption and integration of the FAIR principles.

In summary, while the application of FAIR principles in the financial sector remains limited, evidence from other industries demonstrates their benefits. However, the path toward FAIR-aligned data management must be carefully examined, as there exist challenges in implementing FAIR practices.

3.3 Gaps in the Literature

While the FAIR principles have been widely explored and successfully applied in sectors such as healthcare, agriculture and pharmaceuticals, to our knowledge, their application within the financial sector has not yet been explored. This represents a significant research gap, especially considering the increasing reliance on data-driven decision-making in financial institutions. This research contributes to the field by developing a contextualized FAIR evaluation framework and applying it to traditional banks and P2P lending platforms to assess their alignment with the FAIR principles.

3.4 Summary

This literature review provided insight into the current state of credit risk assessment across both traditional banks and P2P lending platforms, highlighting differences in regulatory requirements, data sources and data management practices. Traditional banks operate within more established legal frameworks and make use of structured, proprietary data management systems. P2P lending platforms, on the other hand, are less bounded by regulations but may lack standardized data processing procedures.

The literature review also examined the application of the FAIR principles. While FAIR-oriented data practices are gaining popularity across domains, their implementation in the financial sector remains limited. However, studies from other disciplines demonstrate the value of applying FAIR principles to improve data management practices. However, transforming data management practices to make them FAIR-compliant also presents challenges that need to be addressed carefully. In the following chapters, these insights are used to develop an evaluation framework to assess the extent to which the credit risk assessment practices of banks and P2P lending platforms are aligned with the FAIR principles.

Chapter 4

Methodology

This chapter presents the methodology used in this research. The research conducts a comparative evaluation to assess the FAIRness of credit risk assessment in traditional banks and P2P lending platforms. Based on differences in data management practices and regulatory frameworks, this research evaluates which financial institution type, banks or P2P lending platforms, more effectively manages and structures its data in alignment with the FAIR principles. This evaluation not only determines which institution more effectively structures its data but also identifies gaps in FAIR data management that could be addressed to improve financial data governance.

To achieve this, this chapter introduces the overall research approach, describes how data is collected and selected and explains how the FAIRness evaluation is applied in practice. This is done through a mixed-methods design, which is used to combine qualitative insights with a structured quantitative assessment based on a predefined evaluation framework.

Section 4.1 describes the research design and methodological steps, Section 4.2 describes the data sources and Section 4.3 the qualitative methods. Together, they form the methodology for the comparative analysis in the subsequent chapters.

4.1 Research Approach

To address the research objectives and evaluate the FAIRness of credit risk assessment in banks and P2P lending platforms, this study adopts a comparative mixed-methods approach [71]. This mixed-methods approach consists of two phases: a qualitative data collection and analysis, and a quantitative data collection and analysis to explain the qualitative data [71]. For the quantitative analyses, a structured evaluation framework based on the FAIR principles is designed.

The overall research process is outlined in Figure 4.1, which presents the key methodological steps. The flowchart provides a visual overview of the structured process used throughout this study. The process begins with the development of a FAIR evaluation framework, which transforms the FAIR principles and sub-principles into assessable criteria. This framework serves as the foundation for the assessment of the FAIR alignment.

After the framework development, a qualitative data collection and analysis is conducted to explore and map the data management practices in traditional banks

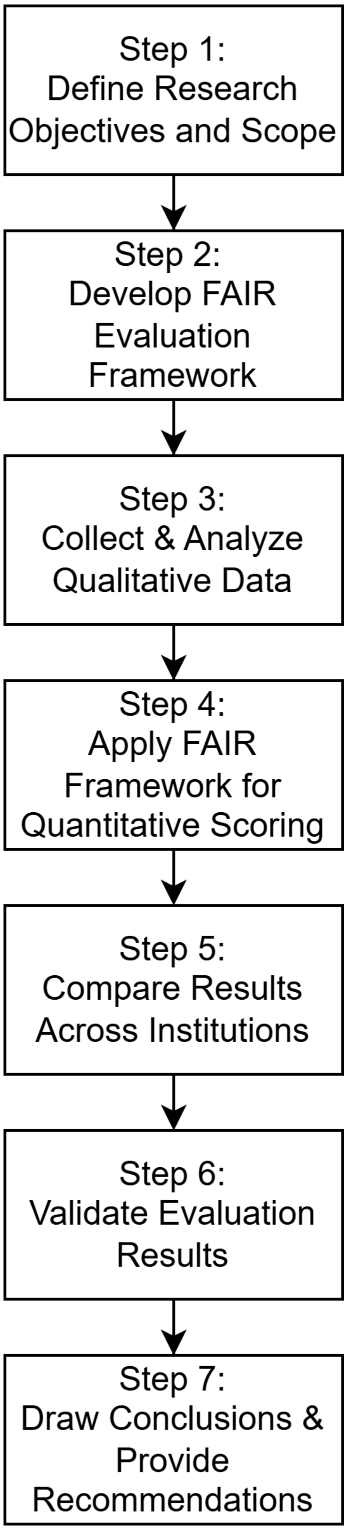

Figure 4.1: Flowchart Methodology

and P2P lending platforms. This phase involves analyzing publicly available institutional reports, regulatory guidelines, academic literature and platform documentation to examine differences in data management and stewardship. The in Chapter 3 conducted literature review forms the basis for this phase. The purpose of this phase is to get a rich understanding of the institutional practices and inform how the FAIR principles are or are not being reflected in practice.

Following the qualitative analysis, a quantitative evaluation is performed for which the previously developed FAIRness framework is used. Based on the findings from the qualitative phase, each institution is scored across the FAIR criteria, enabling a structured comparison of data management practices. This includes assessing the findability, accessibility, interoperability and reusability of the credit risk data using structured metrics and scoring systems to quantitatively measure how well financial institutions align with the FAIR principles. This allows for the identification of which financial institution demonstrates greater adherence to the FAIR principles and in what areas. Further, the scoring also highlights challenges in FAIR alignment, which helps with identifying areas of improvement.

This approach follows a fixed mixed-methods design as described by Creswell and Plano Clark [71], in which a predefined process guides the qualitative and quantitative phases. The overall methodology follows an exploratory sequence, where qualitative data collection is conducted first. This is then subsequently used for the quantitative assessment based on the predefined framework. This structure enables a balanced assessment that captures both the context of the institutions' practices and their measurable alignment with the FAIR principles.

The comparative dimension of the research follows the structured, focused case comparison methodology proposed by George and Bennet [90], which emphasizes the consistent application of evaluation criteria across different institutional cases. This approach supports transparent, reproducible and analytically strict cross-institutional analysis.

4.2 Data Collection and Sources

This research relies primarily on open and publicly available data sources. In addition to the publicly available data, limited non-public input was obtained through a survey. This additional input was used to supplement the publicly available information and enrich the comparative analysis. The data collection process is documented to ensure transparency, reproducibility and consistency of the research. For the two types of institutions that are examined, banks and P2P lending platforms, different sources of data are used to represent their respective data management practices.

All forms of data are collected that give information about data management practices within banks and P2P lending platforms. Regarding the public data this includes institutional documentation such as terms of service, privacy policies and platform documentation. Furthermore, regulatory guidelines published by authoritative bodies such as EBA[1] or BIS[2] are reviewed to understand compliance expectations and data governance standards. Finally, relevant academic literature on credit risk assessment, financial data management and the FAIR principles is consulted to provide theoretical context. As mentioned, this publicly available documentation and literature are supplemented by survey answers.

[1]https://www.eba.europa.eu/homepage

[2]https://www.bis.org/index.htm

The combination of all these data sources is used in the qualitative analysis to map data management practices of the financial institutions. This includes among others data documentation, metadata availability, sharing mechanisms and data origin descriptions. These gathered insights are then used for evaluating the alignment with the FAIR principles with the use of the defined framework.

4.3 Qualitative Methods

This research uses qualitative methods to explore the data management practices of traditional banks and P2P lending platforms in the context of credit risk assessment. The qualitative analysis was primarily based on document analysis, including institutional documentation, regulations and academic literature.

In addition to this document-based analysis, some data was collected through structured survey responses from one representative of a traditional financial institution, such as banks, and one representative from an alternative financial institution, such as P2P lending platforms. The survey is the FAIR alignment assessment, structured in a survey format. The responses offered insight into the actual FAIR alignment within a financial institution. Although the survey results are not generalizable, they provided valuable complementary perspectives that helped verify findings drawn from public sources.

Together, these qualitative inputs were used to develop a detailed understanding of banks' and P2P lending platforms' data management practices.

4.4 Summary

This chapter outlined the methodological foundation of the research. A comparative mixed-methods approach was adopted, integrating qualitative and quantitative methods to enable a comprehensive analysis. The research involved analyzing publicly available institutional, regulatory and academic documentation. Additional input was obtained through structured survey responses from one traditional financial institution and one alternative financial institution. These responses were used to validate the findings from the publicly available data. Combined, these methods enabled a systematic and robust assessment of FAIR alignment in credit risk assessment practices.

Chapter 5

The FAIR Alignment Evaluation Framework

This chapter presents the FAIR Alignment Framework designed to assess data practices in the context of credit risk assessment. The framework translates the FAIR principles into a structured set of evaluation criteria that reflect the specific data practices and requirements of financial institutions. Its aim is to evaluate the extent to which the data, metadata and technical infrastructure used by the financial institutions align with the FAIR principles.

This framework builds on the FAIR Data Maturity Model from the FAIR Data Maturity Model Working Group [84]. Their model presents a set of general indicators based on the FAIR principles, each with an assigned priority level. However, they acknowledge the limitation that FAIR practices can vary significantly between domains. Wilkinson et al. [171] share a similar perspective, they noted that metrics should ultimately be supplemented through the development of domain- or community-specific metrics. To support this, they proposed a metric creation template to guide such efforts. Building on these ideas, the framework presented here is specifically tailored to evaluate the FAIRness of data practices of credit risk assessment within financial institutions.

The FAIR principles themselves are intentionally designed as high-level and domain independent guidelines [169]. Therefore, the principles require contextual interpretation in order to be meaningfully applied. This framework incorporates such contextualization by defining how each principle applies to data governance in the financial sector. Furthermore, the relative importance of individual principles may vary between domains and therefore also needs to be included in the framework.

This chapter begins by defining the scope and targets of the assessment (Section 5.1), followed by the domain-specific conceptualization of the FAIR principles (Section 5.2), an evaluation of their relative relevance (Section 5.3) and finally, the operational structure of the scoring framework itself (Section 5.4). Section 5.5 concludes by highlighting the framework's adaptability and its openness to expert input.

5.1 Scope and Targets

A meaningful evaluation of FAIR alignment begins with a clear definition of what is being assessed. This section outlines the scope of the evaluation by identifying the core components that the framework focuses on.

In line with the framework of the FAIR Data Maturity Working Group [84], which distinguishes between data and metadata as separate assessment dimensions, the FAIRness evaluation in this research focuses on three components:

- Data: the structured information used in credit risk assessment, such as loan attributes, borrower profiles and repayment histories.

- Metadata: the descriptive information that provides context about the data. This includes documentation that helps users and systems interpret the datasets correctly, such as field definitions, data types and version histories.

- Infrastructure: the technical systems that support data storage, access, sharing and reuse. This includes APIs, cloud storage platforms, access control mechanisms, etc. While the framework does not perform a full technical audit, it assesses the presence and functionality of infrastructure components where they impact FAIR related criteria.

By including these three layers in the scope of evaluation, the framework aims to provide a holistic view of the FAIRness level within the financial institutions. Data quality cannot be assessed in isolation from the metadata that describes it or the technical infrastructure that enables the use of the data. They are all necessary to make a full assessment of the alignment with the FAIR principles.

While the framework was only applied to traditional banks and P2P lending platforms, it was designed with broader applicability in mind. The framework can be adapted and reused for FAIR assessments in other financial institutions.

To ensure accuracy and comparability, the framework assesses data and metadata separately, where the respective FAIR principle applies to both. This distinction allows for a more targeted assessment.

5.2 Contextualization of the FAIR principles

As mentioned in the introduction the FAIR principles are intentionally designed as high-level and domain independent guidelines, making them broadly applicable but not immediately operational in specific contexts. Broad terms like "rich metadata" require interpretation. What qualifies as "rich metadata" in one domain may not be sufficient in another domain.

To meaningfully apply the FAIR principles in credit risk assessment, they need to be interpreted in relation to the regulations, internal processes and technical systems of financial institutions. This section provides such domain specific interpretations for each FAIR sub-principle. The goal is not to redefine the principles, but to contextualize them by providing a more detailed explanation of their meaning and implications within the financial domain.

Each FAIR sub-principle is presented with its own short explanation on how it should be interpreted and applied within the context of credit risk assessment in financial institutions. These interpretations are grounded on what is practically feasible within financial institutions, taking into account constraints such as data privacy, governance policies and regulatory compliance.

Findable

F1: (meta)data are assigned a globally unique and persistent identifier.
Every data entity that is key for the credit risk assessment, such as loans, customers and accounts, should have a unique identifier. The Legal Entity Identifier (LEI) is one example of a globally unique identifier which can be used for identifying borrowers across institutions. However, in most cases it is sufficient for identifiers to be unique within the financial institution itself. Regulatory frameworks such as the BCBS 239 emphasize this importance of "single identifiers and/or unified naming conventions" within the financial institution itself [53].

Regarding the persistency of the identifier, this is dependent on the life cycle of the data and regulatory retention policies. Identifiers must remain stable for at least the duration of the data's operational use and retention period, which commonly ranges from ten to twenty years [33]. While this level of stability is the minimum required, permanent functionality, as adopted in the scientific domain, represents a more robust best practice [94].

F2: data are described with rich metadata (defined by R1 below).
All stored data should have extensive metadata, which allows for the data to be easily found again through attribute search [89]. This includes among others field names or labels, definitions, data types, versioning, allowable values, etc. [89]. General metadata standards such as the Dublin Core provide a foundational structure [166], while domain specific frameworks like INEXDA's metadata schema offer more tailored approaches for financial data [56]. This level of detail ensures that humans and machines interpret the data correctly and makes it easier to find the data.

To contextualize richness in this domain, the following metadata categories are considered relevant according to Cedar Rose [22] for credit risk data:

- Descriptive Metadata: Industry classification, legal entity registration, reporting period and dataset title or description.
- Structural Metadata: Includes information hierarchy.
- Legal Metadata: Copyright, licensing terms, regulatory compliance and terms of use.
- Reference Metadata: Definitions of variables and indicators, sources of input data and credit scoring methodologies.
- Statistical Metadata: Credit score calculation methods, risk assessment models and industry benchmarks and comparisons.
- Administrative Metadata: Timestamps of data creation and updates, update frequency, access permissions, and change history.

F3: metadata clearly and explicitly include the identifier of the data it describes.
Including the identifier of the data within its metadata makes sure that they are

linked. This can be achieved by including a unique reference to the data object, such as a dataset identifier, in the metadata. For example, if a loan dataset has an identifier LOANS_2024Q4 in a data inventory, the metadata should clearly state that it describes the dataset with identifier LOANS_2024Q4. This avoids any ambiguity about which data a given metadata file refers to and allows metadata to be stored separately.

F4: (meta)data are registered or indexed in a searchable resource.
In credit risk assessment, data and metadata should be registered in systems that support structured search and retrieval. This can be done, for example, via institutional data catalogs or metadata registries that index datasets and their associated documentation. Users should be able to find relevant resources via searchable attributes such as dataset name, reporting period or data type. For example, a search for "loans 2024 quartile 4" should return all related datasets and their metadata. This principle applies to both data and metadata.

Accessible
A1: (meta)data are retrievable by their identifier using a standardized communications protocol.
(Meta)data should be accessible to authorized users through standardized technical protocols once their identifier is known. The communication protocol should be standardized within the organization. A standardized protocol to retrieve certain data based on the identifier could be an FTP [13]. The key is that the protocol is standardized and well-documented, so that authorized users can retrieve the data via a common method given the data's identifier [13].

A1.1: the protocol is open, free, and universally implementable.
Open protocols refer to the use of industry standard data exchange methods that do not require proprietary or specialized software for basic access. Examples are HTTPS, FTP and REST APIs, which allow users to retrieve data using widely supported technologies [13]. To fulfill this principle, the protocol used should be freely available, well-documented and compatible with commonly used technologies [13].

A1.2: the protocol allows for an authentication and authorization procedure, where necessary.
Financial institutions make use of sensitive data (borrower profiles, repayment history, income data, etc.) for the credit risk assessment, making access control critical [127]. This means that the method used for retrieving data should be capable of restricting access only to authorized users or systems [143]. Within a financial institution, this might mean some form of role-based access control on data [145], where users authenticate using institutional credentials and are granted access based on predefined roles or permissions. The protocols used, as defined in A1.1, should support such controls [13]. Authentication procedures must also comply with relevant legal and regulatory requirements, such as the GDPR.

A2: metadata are accessible, even when the data are no longer available.

Data used in credit risk assessment may be removed over time due to data retention policies, regulatory obligations (such as the GDPR's right to be forgotten) or operational considerations. This principle requires that metadata remain accessible even when the underlying data is no longer available [13]. In practice, this can be implemented by maintaining durable metadata repositories such as enterprise data catalogs [115]. Institutions may annotate such metadata to indicate that the associated dataset has been deleted. Standards such as the Preservation Metadata: Implementation Strategies (PREMIS) can support this process by providing a core set of metadata elements related to metadata preservation [42]. Metadata persistence is not subject to the same retention limits as data and should be maintained indefinitely, unless legal or regulatory constraints dictate otherwise. This principle applies only to metadata.

Interoperable
I1: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
With the knowledge representation language is meant the format and structure in which data and metadata are presented [100]. This should be a format and structure that is both human and machine readable within the financial domain. A "formal" language might be a common data modeling standard or syntax. "Accessible and shared" means that the language is not proprietary and is well-documented, so anyone can learn it. "Broadly Applicable" indicates a language that is widely used across the financial domain [13]. Practical examples include relational database schemas based on the SQL standard, or data interchange formats such as JSON, XML or CSV [109, 130]. This also includes industry ontologies, such as the Financial Industry Business Ontology (FIBO) [57]. Standards like ISO 20022 or XBRL also exemplify formal, shared languages that enable consistent data interpretation across systems [70, 163]. In the absence of standards, one department might code a default as "DEF" in a text field and another as "0" in a numeric field. This principle applies to both data and metadata.

I2: (meta)data use vocabularies that follow FAIR principles.
This principle refers to the use of standardized vocabularies that are themselves FAIR [13]. It is not sufficient to simply standardize field names internally, the vocabulary must be among others well-documented, have a unique and persistent identifier and be described with a formal language so that humans and machines can find, access, interoperate and reuse it [13]. For example, field names should align with widely recognized vocabularies, such as those defined in the Basel Framework glossaries [27]. Common field names like "LGD" (Loss Given Default) should follow established definitions and not rely on institution specific interpretations. Where it is necessary to use internal vocabularies, such as in internal credit risk models, it should still be clearly documented and possibly linked to external standards. This principle applies to both data and metadata when vocabularies are used.

I3: (meta)data include qualified references to other (meta)data.
This principle refers to the use of structured, meaningful links between datasets or

data elements [13]. This includes referencing external data sources (e.g., credit bureaus for the obtained credit scores), linking internal datasets (e.g., customer data linked to loan records) and associating data with its source documentation, such as risk models or regulatory reports. The references should be explicit and machine readable [13]. This principle applies to both data and metadata.

Reusable

R1: meta(data) are richly described with a plurality of accurate and relevant attributes.
This principle is fulfilled when (meta)data include sufficient descriptive detail to support correct reuse beyond the original purpose. This includes among others business context, purpose and limitations [13]. Much of the required richness for metadata was previously outlined in F2. In the context of R1, the focus extends to additional descriptive attributes that enable reuse of both metadata and data.

R1.1: (meta)data are released with a clear and accessible data usage license.
This principle is fulfilled when the conditions under which data and metadata can be accessed, shared and reused are clearly defined and documented. In the context of public research data, R1.1 typically refers to attaching a formal license such as Creative Commons [13]. This form of licensing is uncommon in the domain of credit risk data, as such data is not released publicly. Therefore, R1.1 can be interpreted as ensuring that usage rights and restrictions are clearly stated for whoever is going to use the data. Internal data is usually managed by data classifications (confidential, internal use only, public, etc.) [147]. These conditions should be explicitly documented alongside the data. Legal frameworks such as the GDPR also act as implicit usage constraints. GDPR requires that personal data may only be used for the purposes for which it was collected, which essentially acts as a legal usage license [24]. This serves the same function as licenses by defining how data and metadata may be used, by whom and under what conditions.

R1.2: (meta)data are associated with detailed provenance.
The provenance data should include where the data originated from, who created it, why it was created, how it was created, which version of the data it represents and links to related publications [150]. In the context of credit risk data, this could include identifying a credit bureau as the original data source, recording any data cleaning steps applied and tracking changes in the structure or content of the data over time. This principle applies to both data and metadata.

R1.3: (meta)data meet domain-relevant community standards.
This principle is fulfilled when (meta)data follow established standards that are recognized and used within the financial sector. In the context of credit risk assessment, this includes aligning data management practices with regulatory frameworks such as the Basel Accords, IFRS standards, ISO standards, FIBO and with industry data models such as INEXDA and XBRL. Compliance with this principle requires that both externally reported and internally managed data align with these standards where applicable. This applies to both data and metadata.

5.3 Relevance of FAIR Principles in the Context of Credit Risk Assessment

While all FAIR sub-principles contribute to improving data management, they do not contribute all equally in every domain [170]. Where one sub-principle may be critical for the credit risk assessment process, others might be desirable but less impactful. The relevance of a specific sub-principle depends on the domain in which they are applied. This results in some sub-principles being of higher importance to credit risk assessment than others.

To reflect these differences, each FAIR sub-principle is assigned a relevance score on a three-point scale:

- 1 – Low relevance: The principle has limited applicability to core credit risk practices.
- 2 – Moderate relevance: The principle contributes meaningfully to data management but is not essential in all cases.
- 3 – High relevance: The principle is essential and directly supports core credit risk activities and regulatory obligations.

These scores are based on our own interpretation, grounded in literature and regulatory frameworks. They are intended as a starting point for identifying which principles deserve the most attention in FAIR evaluations. The intention is not to establish a universal ranking, but to indicate which principles are likely to carry the most weight within this particular domain. These assumptions can be further validated or adjusted through consultation with experts.

The following sections present the relevance scores for each sub-principle, together with a brief rationale explaining their relevance in the credit risk assessment.

Findable
F1: (meta)data are assigned a globally unique and persistent identifier.
Relevance: 3
Unambiguous identification is fundamental for credit risk assessment in financial institutions [123]. Without unique identifiers for loans or customers it becomes difficult to reliably connect data and avoid duplication. For example, not recognizing that two loan records refer to the same customer could lead to underestimating the customer's total debt. Regulatory frameworks like BCBS 239 emphasize the need for unified identifiers within institutions, reinforcing the importance of this principle in the domain [53].

F2: data are described with rich metadata (defined by R1 below).
Relevance: 3
Rich metadata is of high importance in the context of credit risk assessment. When metadata is not richly described it not only makes it more difficult to find the data, but it can also lead to misinformed decision-making [78]. In terms of findability,

well-documented metadata improves data discovery, makes linking data easier and avoids unnecessary duplication [176]. Without rich metadata, even well-identified data (F1) could be misinterpreted or remain underutilized, undermining the quality of the credit risk assessment.

F3: metadata clearly and explicitly include the identifier of the data it describes.
Relevance: 3
This principle prevents ambiguity by linking metadata to the specific data it describes [105]. The principle helps to prevent documentation mix-ups, such as applying the wrong definition to a similarly named variable used in a different system [155]. A lack of explicit linkage between metadata and data can significantly slow down data discovery [120]. Users may waste time verifying whether documentation applies to a particular dataset, especially when similar variable names are used across different systems [155].

F4: (meta)data are registered or indexed in a searchable resource.
Relevance: 3
Having a searchable catalog for data is a significant enabler of efficiency and data discovery [132]. As more data sources become available, the importance of a searchable data catalog grows. This is reflected in the common estimate that data scientists spend up to 80% of their time understanding and cleaning poorly documented or hard to find data [77]. A searchable resource not only improves the data discovery it also improves the transparency, accessibility, data quality and increase usage [115]. Furthermore, by registering data in a searchable registry, financial institutions reduce duplications and ensure that data is not siloed [141]. Maintaining a searchable registry also supports compliance with regulatory frameworks [53].

Accessible
A1: (meta)data are retrievable by their identifier using a standardized communications protocol.
Relevance: 3
This principle is highly relevant in the context of credit risk assessment, as the use of a standardized communications protocol impacts the efficiency, reliability, consistency and security of data operations [80]. No matter how well-defined or findable the data is, if the data cannot be retrieved efficiently, it is practically unusable. Standardized communication protocols address this by enabling repeatable, consistent and secure data access [80]. This guarantees that users and systems retrieve the same data in the same way [80]. Without standardized access, different departments may rely on different methods. One might query a central database, while another uses a locally stored file, leading to discrepancies and version conflicts. Further, the EBA guidelines require that credit risk data be accessible without undue delay and with minimal reliance on manual processes [81].

A1.1: the protocol is open, free, and universally implementable.
Relevance: 1
This principle is of relatively low relevance in the context of credit risk assessment.

Financial institutions often use closed proprietary systems to access and manage data [67]. While the use of open, free and universally implementable protocols can support interoperability and trust, it is not a necessity for the credit risk assessment of financial institutions [67]. It will not affect the accuracy of the credit risk assessment. What is important to credit risk assessment is that data is retrievable (A1) and securely accessible (A1.2). The adoption of open standards has mainly been driven by external regulatory requirements rather than internal credit risk assessment needs [133]. The main risk of not adhering to A1.1 lies in long-term inflexibility. For example, dependence on a specific vendor or outdated systems which may hinder data accessibility if support ends. Therefore, while A1.1 reflects good IT practice, its direct impact on the quality of credit risk assessment is limited.

A1.2: the protocol allows for an authentication and authorization procedure, where necessary.
Relevance: 3
Proper authentication and authorization procedures are necessary for safeguarding the sensitive data within financial institutions, like credit risk data, which includes highly sensitive information such as personal income, credit histories and default records [102]. Inadequate authentication/authorization of credit risk data can lead to severe consequences, including financial loss, data breaches, compliance violations and loss of trust [102]. Regulatory frameworks such as the GDPR mandate strict access control policies. Failure to comply with these policies can result in significant financial penalties [24]. Restricting access to authorized users not only preserves the confidentiality of credit data, but also ensures that analytical outputs remain trustworthy [58, 143]. A1.2 directly supports the secure and responsible handling of financial data.

A2: metadata are accessible, even when the data are no longer available.
Relevance: 1
While retaining metadata after the deletion of data aligns with good data governance practices, its direct impact on credit risk assessment is minimal. The credit risk assessment processes typically focus on available data, once the data is removed its associated metadata is rarely needed for operational modeling or analysis. While accessible metadata for outdated data can support reliability and integrity [172], it is not essential to the core credit risk assessment process in financial institutions.

Interoperable
I1: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
Relevance: 2
Using standardized and structured data formats facilitates data integration across systems, departments and external stakeholders [88]. Regulators increasingly require the use of standardized formats, making standardization often unavoidable [76]. The use of a standardized language also significantly enhances the efficient use of data by reducing (meta)data uncertainties, minimizing the need for data transformation, mitigating issues related to missing data and improving the quality of

what one can learn from data [88]. While adherence to formal and broadly applicable languages improves data consistency and reduces translation effort, it is not a strict requirement for conducting effective credit risk assessments. The relevance of this principle depends on the complexity of the institution's data environment. As data environments grow more complex, the importance of a shared data language becomes more relevant [144].

I2: (meta)data use vocabularies that follow FAIR principles.
Relevance: 1
While the use of standardized vocabularies is certainly beneficial for clarity and interoperability [153], its practical relevance to credit risk assessment processes is relatively limited. In many cases, financial institutions already comply with I2 due to regulations or market practice, such as the standard definitions defined by the Basel framework. In situations where standard vocabularies are not used, the main consequence is often the need for mapping or translation when sharing data [13]. A manageable inefficiency rather than a critical risk. Although regulators increasingly promote the use of standardized vocabularies to enhance transparency [146], the absence of FAIR-aligned vocabularies rarely hinders the core processes of credit risk assessment.

I3: (meta)data include qualified references to other (meta)data.
Relevance: 3
Using qualified references to link related data enhances data exploration, facilitates discovery and improves data insights [60]. Regulatory bodies, such as the EBA, also require such linkages between related data, for example loan level information should be linked to related borrower and collateral data [81]. Without proper linkage between related data, a financial institution may fail to recognize critical relationships. For example, it might not get recognized that several loans are tied to the same customer, potentially underestimating total exposure or miscalculating default probabilities. To prevent this from happening proper references should be in place.

Reusable
R1: meta(data) are richly described with a plurality of accurate and relevant attributes.
Relevance: 3
Richly described metadata is critically important, as it enhances overall data quality, supports validation processes, facilitates regulatory compliance and significantly strengthens data accessibility and usability [37]. If data lacks some contextual details, reusing the data can lead to misinterpretation of the data [37]. Although context can sometimes be manually inferred, this is inefficient and error-prone. As data environments grow larger and more complex, extensive metadata is essential to ensure correct and reliable reuse of data [37]. R1 is therefore essential for scalable and reliable credit risk analysis.

R1.1: (meta)data are released with a clear and accessible data usage license.
Relevance: 3

If this principle is interpreted as ensuring that usage rights and restrictions are clearly stated for whoever is going to use the data, instead of requiring a formal license, then this principle is of high relevance. Within a financial institution, credit risk data is often restricted based on access level or sensitivity. The absence of clearly defined usage conditions can result in inappropriate data sharing or misuse, leading to legal and reputational risk [136]. Additionally, legal frameworks like the GDPR impose strict requirements on the use of personal data, which essentially acts as a legal usage license. Clear documentation of usage rights makes compliance with permitted use easier. The absence of such clarity increases the risk of unauthorized access, misuse or legal breaches [136].

R1.2: (meta)data are associated with detailed provenance.
Relevance: 3
Detailed provenance enhances data quality and reliability, supports audit trails and error detection, enables the replication and updating of data processes, establishes attribution and ownership, and provides essential context for data discovery and interpretation [150]. For example, if a probability of default (PD) seems off, provenance records can reveal whether input data was manually altered or affected by a system issue. Without detailed provenance, the credibility of credit risk reports and models is undermined and institutions may be unable to justify their decisions to auditors or supervisors. Additionally, it can lead to undetected model errors. Effective recordkeeping of the (meta)data is important to ensure the reliability, authenticity, usability and integrity [172].

R1.3: (meta)data meet domain-relevant community standards.
Relevance: 2
Many of these standards are mandated by regulators [146]. Hence, credit risk data used for supervisory reporting is already aligned with community standards by requirement. Adhering to these standards makes it easier to accurately and efficiently share information [50]. In the absence of such standards, the data must undergo additional processing or translation to be usable [50]. However, as long as the data can be translated to be usable it is not critical to use community standards internally. R1.3 provides clear benefits, but is not absolutely critical, provided the institution can translate the data to standard forms when required.

5.4 Scoring and Operationalization of the FAIRness Framework

This section presents the operational structure of the FAIR alignment framework, which translates each FAIR sub-principle into measurable evaluation criteria. The aim is to provide a systematic and transparent method for assessing the extent to which financial institutions' credit risk data practices align with the FAIR principles. As mentioned in the introduction this idea builds upon the FAIR Data Maturity Model developed by the FAIR Data Maturity Model Working Group [84] and Wilkinson et al. guidance on metrics [171].

The framework is structured around three components: targets (e.g., data, metadata, infrastructure), evaluation aspects (specific features being assessed) and criteria (observable conditions). Each sub-principle is broken down into one or more criteria, allowing for detailed evaluation across different systems or organizations. In total, the framework comprises the 15 FAIR sub-principles, which are assessed using 49 evaluation criteria.

To ensure comparability, each principle is normalized to contribute equally to the overall FAIR alignment score. The total score for a sub-principle is the sum of its fulfilled criteria, divided by the number of criteria it contains. This ensures that each FAIR sub-principle contributes equally (maximum of 1 point) to the overall alignment score, regardless of how many criteria it includes. Institutions may adjust the criteria, modify thresholds or even add or remove evaluation aspects depending on their specific data environment and regulatory context. Any adjustments must maintain the equal weighting of the principles for comparability.

The remainder of this section details the scoring tables for each FAIR sub-principle. The criteria for sub-principles F1, A1, A1.2, I1, and R1.3 are primarily inspired by the FAIR Data Maturity Model [84], while the remaining criteria are formulated to comprehensively reflect the intent of each FAIR sub-principle.

TABLE 5.1: Evaluation Criteria for Findability

ID	**Target**	**Evaluation Aspect**	**Criterion**
F1: (meta)data are assigned a globally unique and persistent identifier.			
F1-1	Data	Global Uniqueness	Each data object is assigned an identifier that is unique at the institutional level. Where possible, globally unique identifiers, like the Legal Entity Identifier (LEI), are used.
F1-2	Data	Persistence	Identifiers remain stable and unchanged throughout the data lifecycle, this includes among others system migrations, updates and archiving.
F1-3	Metadata	Global Uniqueness	Each metadata record is assigned an identifier that is unique within the institution.
F1-4	Metadata	Persistence	Identifiers remain stable and unchanged throughout the data lifecycle, this includes among others system migrations, updates and archiving.
F2: data are described with rich metadata (defined by R1 below).			
F2-1	Metadata	Descriptive Attributes	Metadata includes core descriptive elements that support effective search and discovery, such as title, description, keywords, format and classification terms.

F2-2	Metadata	Provenance Information	Metadata includes basic descriptive elements such as title, description, keywords, format and contextual notes (e.g., calculation methods).
F3: metadata clearly and explicitly include the identifier of the data it describes.			
F3-1	Metadata	Data Linkage	Metadata includes an explicit reference to the corresponding dataset using a unique and stable identifier.
F3-2	Metadata	Machine Readability	The linkage between metadata and data is represented in a structured, machine-readable format.
F4: (meta)data are registered or indexed in a searchable resource.			
F4-1	Data	Registration/ Indexing	Data is registered/indexed in a searchable resource.
F4-2	Metadata	Registration/ Indexing	Metadata is registered/indexed in a searchable resource.
F4-3	Infrastructure	Search Functionality	The searchable resource supports keyword or attribute-based search across datasets and metadata records.
F4-4	Infrastructure	Coverage	The searchable resource provides comprehensive coverage of all relevant datasets and metadata.
F4-5	Infrastructure	Machine Readability	The searchable resource provides a machine-readable interface for automated search and discovery.

Table 5.2: Evaluation Criteria for Accessibility

ID	**Target**	**Evaluation Aspect**	**Criterion**
A1: (meta)data are retrievable by their identifier using a standardized communications protocol.			
A1-1	Data	Identifier-Based Retrieval	Data can be retrieved using their identifier via a communication protocol (e.g., HTTPS or FTP).
A1-2	Metadata	Identifier-Based Retrieval	Metadata is retrievable using its identifier via the same communication protocol as the associated data.
A1-3	Infrastructure	Standardized Protocol	The communication protocol used to access (meta)data is well-documented and based on established standards.
A1.1: the protocol is open, free, and universally implementable.			

A1.1-1	Infrastructure	Use of Open and Free Protocols	Data retrieval relies on open, non-proprietary communication protocols that are freely and universally implementable.
A1.1-2	Infrastructure	Protocol Documentation	The communication protocols used are publicly documented and accessible.
A1.1-3	Infrastructure	Consistent Implementation	The same protocol is consistently applied across systems and departments within the institution.
A1.1-4	Infrastructure	Automation Support	The protocol allows for automated data retrieval without requiring manual intervention.
A1.2: the protocol allows for an authentication and authorization procedure, where necessary.			
A1.2-1	Infrastructure	Authentication Support	The communication protocol supports authentication to verify the identity of users or systems accessing (meta)data.
A1.2-2	Infrastructure	Authorization Control	The protocol supports authorization mechanisms to control access to (meta)data.
A1.2-3	Infrastructure	Regulatory Compliance	Access control mechanisms comply with relevant legal, regulatory and institutional requirements.
A2: metadata are accessible, even when the data are no longer available.			
A2-1	Metadata	Metadata Availability	Metadata remains accessible through the same infrastructure, even after the associated dataset is archived or deleted.
A2-2	Metadata	System Support	Metadata is stored in a durable and searchable system designed for long-term access, independent of the related data's lifecycle.
A2-3	Metadata	Metadata Annotation	Metadata records indicate when the associated data is no longer available and include a reason or status (e.g., archived, deleted).

Table 5.3: Evaluation Criteria for Interoperability

ID	**Target**	**Evaluation Aspect**	**Criterion**
I1: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.			

I1-1	Data & Metadata	Use of Structured Formats	Meta(data) is stored and exchanged in structured, machine-readable formats (e.g., JSON, XML, CSV) that support human and system interpretation.
I1-2	Metadata	Use of Standardized Schemas	Metadata follows standardized and widely adopted schemas (e.g., DDI) where applicable.
I1-3	Infrastructure	Language Accessibility	The data and metadata formats used are open, publicly documented and commonly adopted across the financial domain.
I2: (meta)data use vocabularies that follow FAIR principles.			
I2-1	Data & Metadata	Use of Standardized Terms	Variable names, classifications and terminology used in (meta)data conform to recognized domain vocabularies.
I2-2	Metadata	Vocabulary Documentation	Vocabularies are documented, versioned and accessible to data users for interpretation and integration.
I2-3	Metadata	FAIR Vocabulary Alignment	Vocabularies are themselves FAIR: they use a unique and persistent identifier, are resolvable via a standardized protocol and use a formal knowledge representation language.
I2-4	Metadata	Machine-Readability	Vocabularies are available in a machine-readable format.
I3: (meta)data include qualified references to other (meta)data.			
I3-1	Data	Referencing	Data includes explicit references to other datasets or tables, using unique keys or identifiers that support linkage.
I3-2	Metadata	Reference Clarity	Metadata defines and documents relationships between datasets, including the nature and direction of the linkage.
I3-3	Infrastructure	Machine-Readable Linking	References are stored in structured, machine-readable formats.

Table 5.4: Evaluation Criteria for Reusability

ID	**Target**	**Evaluation Aspect**	**Criterion**
R1: meta(data) are richly described with a plurality of accurate and relevant attributes.			
R1-1	Metadata	Technical Information	Metadata includes essential technical attributes such as definitions, data types, units of measurement and timestamps.

R1-2	Metadata	Contextual Information	Metadata captures contextual information such as the purpose of data collection, applicability and known limitations.
R1-3	Metadata	Versioning	Metadata includes version information and a documented change history to track modifications over time.
R1-4	Metadata	Reuse Support	Metadata is sufficiently detailed to support the correct interpretation and reuse of data in different operational contexts.
R1.1: (meta)data are released with a clear and accessible data usage license.			
R1.1-1	Data & Metadata	Usage Documentation	Clear usage rights and restrictions are stated for both data and metadata.
R1.1-2	Data & Metadata	Legal Compliance	Stated usage rights comply with applicable legal regulations (e.g., GDPR) and institutional data policies.
R1.1-3	Data & Metadata	License Transparency	Usage terms are easy to locate, accessible and written in understandable language.
R1.2: (meta)data are associated with detailed provenance.			
R1.2-1	Metadata	Source Documentation	Metadata describes the data's origin and the conditions under which it was collected.
R1.2-2	Metadata	Transformation Records	Metadata documents changes to the data, including transformations, cleaning or aggregation steps.
R1.2-3	Metadata	Version Control and Timestamps	Metadata includes timestamps, version information and identification of responsible individuals or systems for updates.
R1.2-4	Metadata	Systematic Recordkeeping	Provenance information is systematically maintained as part of the data management process.
R1.3: (meta)data meet domain-relevant community standards.			
R1.3-1	Data & Metadata	Domain-Specific Vocabulary	(Meta)data use terminology that aligns with established community conventions or domain-specific language.
R1.3-2	Metadata	Structured Format Alignment	Metadata follows domain-recognized structures or standards.

5.5 Summary

This chapter has introduced the FAIR alignment framework, a structured and domain specific tool for evaluating the FAIRness of data practices in credit risk assessment. The framework translates the FAIR principles into measurable criteria.

The chapter contextualized each FAIR sub-principle within the domain of credit risk assessment, assessed their relative relevance and transformed them into a measurable criterion based scoring system. The scoring system allows for consistent evaluation, but also remains adaptable to institutional variation.

In the following chapter, this framework will be applied to assess and compare the FAIR alignment of the credit risk assessment within banks and P2P lending platforms, demonstrating its usefulness and adaptability in practice.

Chapter 6

Results and Discussion

This chapter presents and interprets the results of the FAIRness assessments conducted on traditional banks and P2P lending platforms. Based on the detailed scores provided in Appendices A and B, as well as the survey responses in Appendices C and D, the analysis highlights the key differences and similarities across the four FAIR principles: Findability, Accessibility, Interoperability and Reusability. The scoring across these four FAIR principles is summarized in Figure 6.1. By synthesizing the results, the following sections examine how the two types of financial institutions align with the FAIR principles and explore the broader implications for data management practices in the credit risk domain.

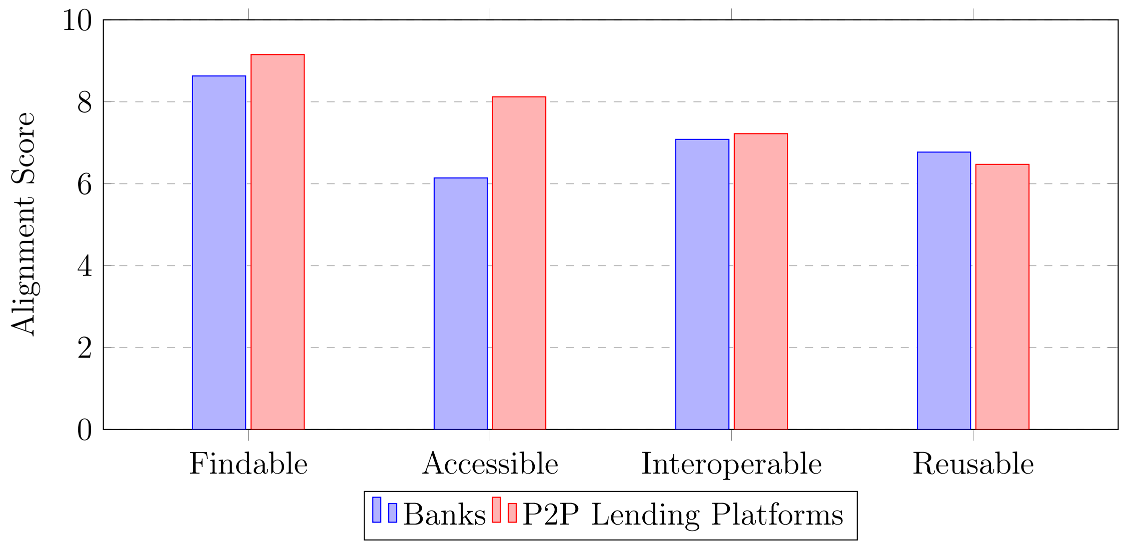

FIGURE 6.1: FAIR alignment comparison of Banks and P2P Lending Platforms based on publicly available information and survey answers

6.1 Findability

The FAIR alignment assessment shows that both traditional banks and P2P lending platforms have good performance in regards with the Findability principle. Both demonstrate good alignment with the first three sub-principles, while scoring slightly lower on criterion F4.

For traditional banks, public data revealed a reliance on regulated frameworks that standardize data structures and reporting formats, including the use of unique and persistent identifiers (Appendix A). However, these publicly available regulatory frameworks do not explicitly prescribe the implementation of structured metadata schemas. To address this, collaborative initiatives such as INEXDA[1] have emerged, which promote best practices for data documentation among participating institutions. Additionally, many banks use some form of data cataloging tool, this can be provided by an external software company such as Collibra[2], Alation[3] or Informatica[4] or developed internally. Such data cataloging tools support attribute-based search, automatic discovery and indexing of metadata. The use of such a cataloging tool was confirmed in the survey response (Appendix C), which also verified consistent metadata-to-data linkage and rich metadata.

P2P lending platforms, on the other hand, achieve comparable Findability scores through different approaches. Public documentation shows that these platforms typically expose data through APIs and leverage cloud-based data governance tools that enable stable identifiers and machine-readable formats (Appendix B). These help answer most criteria. However, some criteria require access to internal metadata practices. The public availability of this type of information is limited. To address this gap, the European Single Access Point (ESAP) Regulation was used as a reference point in the assessment. While P2P lending platforms are not currently required to comply with ESAP, the regulation is a new and explicitly forward-looking initiative that anticipates future inclusion of additional financial actors. Given that P2P platforms are already regulated under the Crowdfunding Regulation (EU 2020/1503) and are going to be included in the upcoming Financial Data Access (FIDA) Regulation, they are likely candidates for eventual inclusion in ESAP. Therefore, the ESAP framework is used to answer certain criteria for metadata practices in the assessment through publicly available data. Since ESAP places a strong emphasis on findability for data and metadata requirements, the relevant criteria for this principle are largely met. The survey response (Appendix D) also revealed strong alignment with the Findability principle. The survey showed use of internal identifiers, extensive metadata and the use of a searchable resource with comprehensive coverage of all data, leading to a slightly higher overall Findability score than that of banks.

The findings show an important methodological insight: public documentation may reflect compliance goals, while internal assessments reveal actual operational practices. In both cases, the survey responses provided essential context to interpret the observed FAIRness. This shows the value of combining external and internal evaluation methods. The combination of public and internal sources provided a more balanced evaluation of findability practices.

[1]INEXDA: https://www.inexda.org/
[2]Collibra: https://www.collibra.com/
[3]Alation: https://www.alation.com/
[4]Informatica: https://www.informatica.com/

6.2 Accessibility

For accessibility P2P lending platforms show a significantly better alignment with the FAIR principles than traditional banks. Public documentation indicates that both sectors implement similar technical practices, notably the use of APIs and standardized communication protocols, but these are embedded within different institutional and regulatory frameworks. Further, the survey responses reveal differences in internal practices that are not reflected in publicly available data.

Traditional banks typically provide access to datasets through secure internal infrastructures. In some cases, banks also expose developer-facing APIs. This can be seen in the practices of some major Dutch banks such as ING, ABN Amro and Rabobank (Appendix A). Their developer portals offer APIs and supporting documentation that enable access to selected datasets. These APIs use secure communication protocols such as HTTPS in combination with authentication and authorization mechanisms such as OAuth 2.0. This design enables reliable and secure access to data. These interfaces are designed primarily for partners, regulators and third-party developers and are governed by clearly defined usage and access policies. While this setup aligns with many criteria for Accessibility, the survey results (Appendix C) showed some shortcomings in internal data practices. Specifically, the institution reported a lack of publicly available communications protocols, indicating a reliance on proprietary communication protocols. Further, the survey showed inconsistent use of protocol applications across systems, a lack of integrated authentication and authorization mechanisms and insufficient infrastructure to ensure metadata remains accessible after data has been archived or deleted. While it was reported that metadata can indicate when data is no longer available (A2-3), it also acknowledged that metadata is not stored in durable, searchable systems (A2-2) and is inaccessible once data is deleted (A2-1). This may indicate that such metadata exists in isolated formats, such as manual logs or departmental documentation, rather than as part of a structured metadata management framework. Furthermore, certain technical capabilities were clearly described in public documentation, yet the survey responses indicated that these mechanisms were not implemented internally. This suggests that FAIR aligned practices are isolated to external services and have not been fully integrated into the institution's internal systems.

In contrast to traditional banks, P2P lending platforms often provide broader access to their datasets through publicly available APIs. These interfaces are designed for end users, developers and partners. Public documentation showed that communication protocols are open, non-proprietary, secure and documented (Appendix B). The survey response partially confirmed this view, noting the use of open and secure protocols but also identifying gaps in documentation and implementation consistency (Appendix D). Nevertheless, the alternative financial institution scored higher overall due to fully implemented authentication and authorization mechanisms, as well as better practices for storing metadata in a durable, searchable infrastructure.

The key distinction between the alternative financial institution and the traditional financial institution is the support of authentication, authorization and long-term accessibility of metadata through a structured metadata management framework by the alternative financial institution. Further, the traditional financial institution showed a big difference between external capabilities, which are publicly available, and internal capabilities. The internal alignment with the Accessibility principle was significantly less than the publicly available data showed it was capable of.

6.3 Interoperability

Both traditional banks and P2P lending platforms demonstrate moderate alignment with the Interoperability principle. Public documentation shows that both use structured, machine-readable formats for data exchange, such as XML and JSON. However, the depth of alignment varies across sub-principles, and the survey responses reveal internal differences that are not visible through public documentation alone.

Traditional banks achieve interoperability primarily due to the influence of regulatory frameworks such as the Statistical Data and Metadata exchange (SDMX) framework, which standardizes data structures for reporting purposes. In addition, some banks also participate in initiatives such as INEXDA, which promotes standardized metadata documentation practices (Appendix A). Publicly, the main limitation lies in the absence of evidence that these banks use formal, machine-readable vocabularies that meet the FAIR criteria under I2 (Appendix A). The survey response (Appendix C) confirms that banks do use these controlled vocabularies. Further, the survey response also confirms the use of structured and machine-readable data formats. However, the survey response also indicates limitations such as cross-dataset linkages.

P2P lending platforms also meet basic interoperability requirements mainly through their use of structured, machine-readable formats such as JSON and XML (Appendix B & D). These formats support technical interoperability, but the P2P lending platforms lack publicly documented use of controlled vocabularies and documentation on metadata schemas. The survey response (Appendix D), however, provides a more complete picture. It confirms that the platform uses standardized metadata schemas, relies on documented and versioned vocabularies and explicitly defines relationships between datasets in machine-readable ways. This survey response results in high internal scores across all three Interoperability sub-principles.

Based on a combined scoring of publicly available data and the survey responses, banks and P2P lending platforms have similar scoring. However, this is mainly due to gaps in publicly verifiable information for P2P lending platforms, not necessarily weaker practices. The survey shows that the P2P lending platforms achieve almost a perfect score on Interoperability. The survey fills the gaps in the public analysis of the Interoperability alignment.

6.4 Reusability

Although some individual criteria are met, neither institution demonstrates strong compliance with the full set of requirements. However, both show metadata standardization, usage tracking and contextual enrichment.

Traditional banks benefit from mature metadata management practices driven by regulatory compliance and institutional governance structures. Public documentation (Appendix A) shows that metadata typically includes essential technical attributes such as data types, units and timestamps, based on the existence of metadata schemas, such as INEXDA. The metadata also includes usage terms and is sufficiently detailed to support the correct reuse of the data. The survey response (Appendix C) confirms these practices and adds that provenance data is better maintained than publicly available data shows. However, limitations remain in documenting changes over time and in aligning metadata structures with community standards.

Based on publicly available data, P2P lending platforms score lower on most Reusability sub-principles. This is mainly due to the fact that there is no explicit metadata schema publicly available to represent metadata practices. Based on public information that is available, it is concluded that their metadata is often sparse and lacks detailed contextual information (Appendix B). However, survey results (Appendix D) provide a more complete picture. The platform reports support for technical attributes, contextual metadata and change tracking. This suggests stronger alignment with the Reusability principle than was visible through public sources. A notable limitation is the partial support for provenance tracking.

While more extensive publicly available metadata can be found for traditional banks, the internal assessment shows that P2P lending platforms offer similar alignment with the Reusability principle. This again highlights the importance of integrating internal assessments into FAIR evaluations, as public documentation alone may not perfectly reflect actual operational practices.

6.5 Overall Comparison and Implications

The overall FAIR alignment assessment shows that there are clear differences in the way traditional banks and P2P lending platforms align with the FAIR principles. While initial impressions based solely on public data indicated that traditional banks were more FAIR, survey responses show that P2P lending platforms may actually demonstrate stronger alignment in internal practices for certain FAIR principles.

Both institution types score well on Findability. Traditional banks benefit from regulatory frameworks and participation in metadata standardization efforts like SDMX and INEXDA, which promote the use of structured metadata schemas. This alignment is also reflected internally, with the survey response showing the use of rich metadata, clear data-to-metadata linkages and cataloging tools that support internal findability. P2P lending platforms also show strong Findability through their API infrastructures. Their survey response further confirms the use of internal identifiers and good metadata practices.

The biggest contrast between public documentation and survey responses appears in Accessibility. Public documentation suggests that both banks and P2P lending platforms provide API access using open and secure communication protocols. However, the survey results show that traditional banks rely on proprietary internal communication protocols, have inconsistent protocol usage across departments and lack authentication and authorization mechanisms. In contrast, P2P lending platforms do report the use of open, universally implementable protocols with consistent access control mechanisms and a durable metadata infrastructure similar to what was found in public documentation.

In terms of Interoperability, both sectors perform well, using structured, machine-readable formats like JSON and XML. P2P lending platforms cover almost the full set of principles of Interoperability. Traditional banks also meet most criteria, but lack explicit references linking data to other related datasets.

When it comes to Reusability, both sectors face challenges. Traditional banks show relatively strong contextual metadata and technical attributes, but lack detailed, system-wide provenance tracking and change history. P2P lending platforms report similar strengths. The strong contextual metadata supports the correct interpretation and reuse of data for both. Like banks, P2P lending platforms also show gaps in version tracking and systematic provenance documentation.

Based on public data alone, banks show a better FAIR alignment than P2P lending platforms, this is mainly due to the lack of available public information for P2P lending platforms. A more complete view that includes internal practices actually shows that the FAIR alignment of banks and P2P lending platforms is closer to each other. Notably, P2P lending platforms outperform banks on Accessibility.

In conclusion, the findings indicate that neither traditional banks nor P2P lending platforms can be said to "better align" with the FAIR principles across the board. Instead, each type of institution has specific strengths and weaknesses. Overall, banks and P2P lending platforms are broadly comparable in terms of their current level of FAIR alignment.

6.6 Requirements to Become More FAIR

Achieving better alignment with the FAIR principles requires both technical improvements and broader institutional support. While traditional banks and P2P lending platforms show progress in different areas, neither institution type achieves full FAIR compliance. The following recommendations outline what is needed for a better FAIR alignment.

6.6.1 Banks

Traditional banks already operate within mature regulatory frameworks and often use formalized data standards. However, internal FAIR alignment can be significantly improved by addressing gaps in accessibility, provenance tracking and long-term metadata support. Key recommendations to improve the overall FAIR alignment include:

- **Improving metadata durability**: Although banks use data cataloging tools, survey responses indicate that metadata is not always preserved in a durable infrastructure, especially when datasets are archived or deleted. Institutions should enhance metadata life-cycle management to ensure that records remain accessible and interpretable over time.

- **Improving accessibility infrastructure**: Internal systems are currently based on proprietary and inconsistent data retrieval and lack authentication and authorization mechanisms. Aligning internal systems with the external API capabilities would improve accessibility and even interoperability.
- **Enhancing provenance tracking**: To meet all Reusability criteria, banks should improve the systematic tracking of data origin, transformations and update history across systems.

6.6.2 P2P Lending Platforms

P2P lending platforms generally show stronger alignment with technical FAIR elements. However, they often lack a mature regulatory framework. Key areas for improvement include:

- **Engaging with regulatory initiatives**: As the sector becomes more regulated, P2P lending platforms should proactively adapt to the expected FAIR-related requirements. This would reduce compliance burdens in the future.
- **Strengthening provenance tracking**: Although P2P lending platforms already demonstrate partial support for data lineage and versioning, full alignment with Reusability requires systematic documentation of data origin, transformation steps and update responsibility. Implementing a standardized provenance model would improve this.
- **Enabling machine-readable access and discovery**: While APIs and metadata systems are in place, the platform currently lacks support for fully automated, machine-actionable discovery and retrieval. To address this, interfaces need to be made available that not only serve human users but also support structured, automated queries by systems.

6.7 Managerial Implications

The findings of this research have a number of important implications for data management, compliance strategies and platform development within both traditional banks and P2P lending platforms. While the technical FAIR assessment identifies specific areas for improvement, the broader business and operational consequences are equally important.

For banks, the findings show that while external systems, such as APIs used for open banking and reporting, are often advanced, internal practices lag behind. Metadata accessibility, consistent protocol use and provenance tracking are not fully in place in internal systems. This can lead to problems with data quality and makes it harder for teams to work together. Managers should aim to apply the same data and metadata standards inside the bank as they do for external systems. This would make it easier to find and share data, reduce duplicate work and make reporting to regulators more efficient.

P2P lending platforms show stronger alignment with technical aspects of FAIR, particularly in terms of structured data formats and secure API interfaces. However, these platforms often lack formal documentation and consistent provenance tracking. As new regulations are introduced, these platforms will be expected to meet higher standards. Managers should prepare by improving how they document data, track changes and make data easier to find and use by other systems. These steps would not only make compliance easier but also help build trust with partners and customers.

In both types of institutions, improving FAIR alignment is not just a technical task, it also supports better business decisions, lower risks and more efficient operations. Managers who recognize this can use the FAIR principles to improve the data infrastructure.

6.8 Limitations

While the FAIRness assessment provides a structured comparison between traditional banks and P2P lending platforms, several limitations should be acknowledged to contextualize the findings and inform future research.

First, the assessment relies on publicly available documentation in combination with survey responses. While the survey responses add some insight into internal practices, the sample size is limited. Only one traditional bank and one alternative financial institution participated in the survey. Their responses may not fully represent the diversity of practices across the broader sectors.

Second, in the case of the assessment P2P lending platforms based on public data, the lack of transparent metadata practices made it necessary to approach certain evaluation criteria using external benchmarks. The most notable one is the ESAP Regulation. Although ESAP offers a forward-looking standard for metadata quality, P2P lending platforms are not currently required to comply with it. Using it as a reference point introduces assumptions that may not align with current practices of those platforms, particularly when it comes to internal metadata practices.

Third, despite the survey responses, the study was conducted without in-depth interviews or direct access to internal documents beyond survey data. As a result, some nuances of institutional data management, technical implementations and policy enforcement might remain uncovered. This could particularly affect understanding of how FAIR principles are applied across different departments or operational contexts within institutions. Additionally, because there was no direct contact or follow-up interview with respondents, there is a chance that some survey questions were misunderstood or interpreted differently than intended, which could affect the accuracy or consistency of the responses.

Finally, the analysis still partially depends on publicly available documentation, such as academic literature, developer portals, regulations and institutional websites. While these sources are valuable for assessing compliance and external transparency, they may not fully capture actual internal operational practices. The observed differences between public documentation and internal survey responses demonstrate this limitation and highlight the importance of combining both external and internal perspectives for future FAIR assessments.

Despite these limitations, the study offers valuable insights into the FAIR alignment of traditional banks and P2P lending platforms and highlights the importance of integrating multiple data sources to achieve a comprehensive assessment.

Chapter 7

Conclusion

This thesis aimed to answer which credit risk assessment institution, traditional banks or P2P lending platforms, better aligns with the FAIR principles. By using a structured, criteria based framework, the study assessed the institutions through publicly available information and a survey. In doing so, it was shown how differences in regulation, technological infrastructure and data governance influence the alignment with the FAIR principles.

Based on both publicly available information and gathered survey responses, the findings show that neither traditional banks nor P2P lending platforms fully outperforms the other across all FAIR principles. Instead, each type of institution has specific strengths and weaknesses. Overall, banks and P2P lending platforms are broadly comparable in terms of their current level of FAIR alignment.

Both types of institutions demonstrate strong alignment with the Findability principle. Traditional banks benefit from regulatory structures and standardized metadata practices, while P2P lending platforms rely on publicly accessible APIs that support discoverability. Survey responses confirm that both institutions maintain robust metadata systems.

The biggest differences can be found in Accessibility. Publicly available documentation suggests similar capabilities for both sectors through APIs and open protocols. However, survey responses reveal that traditional banks rely internally on proprietary communication protocols, with inconsistent implementation, no access controls and limited durable metadata storage. In contrast, P2P lending platforms reported open, standardized protocols and consistent access control practices, similar to the capabilities shown in their public APIs. This indicates higher internal alignment with Accessibility principles for P2P lending platforms.

In Interoperability, both institutions perform well technically, using machine-readable formats like JSON and XML. However, banks fall short in explicitly referencing related datasets, while P2P lending platforms show better coverage of interoperability criteria in survey responses.

For Reusability, both sectors face similar challenges. Both traditional banks and P2P lending platforms show rich metadata that is sufficiently detailed to support correct interpretation and reuse of data, but both lack systematic provenance tracking.

Overall, the results indicate that traditional banks and P2P lending platforms are more comparable in FAIR alignment than initially assumed. While public data alone suggested banks perform better, internal practices, particularly around Accessibility, show that P2P lending platforms may, in some aspects be better. Overall, this study concludes that the two sectors show similar alignment with Findability and Reusability, while P2P lending platforms show better alignment with Accessibility and slightly better alignment with Interoperability.

This research contributes new knowledge by applying the FAIR principles systematically in the financial domain, combining public documentation with internal survey data. It provides the first comparative FAIR assessment between traditional banks and P2P lending platforms, highlighting how regulatory frameworks, technological maturity and governance structures influence FAIR alignment.

7.1 Future Research

Future research could extend this analysis in several ways. First, a larger survey involving more banks and P2P lending platforms would help validate the findings and improve generalizability. Second, qualitative methods such as interviews with data experts could provide richer insight into how FAIR principles are practically implemented and interpreted within institutions. Third, future studies could develop institution-specific FAIR maturity models to provide tailored guidance for financial institutions. Finally, as regulations like ESAP and FIDA evolve, follow-up studies could examine how regulatory changes impact FAIR adoption in alternative financial institutions.

In conclusion, this study demonstrates that while financial institutions are making progress toward FAIR data management, full alignment is still a work in progress. As financial institutions become increasingly data-driven, improving FAIR compliance will be essential for their data management practices.

Bibliography

[1] A Peer-to-Peer Lending Platform Data Model | Vertabelo Database Modeler. URL: https://vertabelo.com/blog/connecting-borrowers-and-lenders-a-peer-to-peer-lending-platform-data-model/.

[2] ABN AMRO | Developer Portal | API Products. URL: https://developer.abnamro.com/api-products.

[3] About ISO 20022 | ISO20022. URL: https://www.iso20022.org/about-iso-20022.

[4] About this privacy statement. URL: https://www.abnamro.nl/en/personal/overabnamro/privacy/about-privacy-statement.html.

[5] AI Powered Cloud Data Management | Informatica. URL: https://www.informatica.com/.

[6] AnaCredit. URL: https://www.ecb.europa.eu/stats/ecb_statistics/anacredit/html/index.en.html.

[7] API Documentation | Rabobank Developer Portal. URL: https://developer.rabobank.nl/index.php/api-documentation.

[8] APIs Overview - APIs. URL: https://developer.ing.com/api-marketplace/marketplace.

[9] Atlan | Third-Gen Data Catalog. URL: https://atlan.com/home/.

[10] Basics of What a Credit Bureau Is and Does, Plus Major Ones. URL: https://www.investopedia.com/terms/c/creditbureau.asp.

[11] Bondora API v1 Documentation. URL: https://api.bondora.com/doc?v=1.

[12] DDI CV Platform. URL: https://rdf-vocabulary.ddialliance.org/.

[13] FAIR Principles - GO FAIR. URL: https://www.go-fair.org/fair-principles/.

[14] GESIS - Leibniz Institute for the Social Sciences.

[15] Guideline for INEXDA Metadata Tool (Version: 07 May 2018) 1. Background. URL: https://www.bundesbank.de/Navigation/EN/Bundesbank/Research/RDSC/rdsc.html.

[16] Guidelines on credit data statistics (AnaCredit) | Deutsche Bundesbank. URL: https://www.bundesbank.de/en/service/reporting-systems/guidelines-on-credit-data-statistics-anacredit--803824.

[17] How Banking Works, Types of Banks, and How To Choose the Best Bank for You. URL: https://www.investopedia.com/terms/b/bank.asp.

[18] How is Bondora regulated? – Bondora Help Center. URL: https://help.bondora.com/hc/en-us/articles/14898302812689-How-is-Bondora-regulated.

[19] ISO - International Organization for Standardization. URL: https://www.iso.org/home.html.

[20] ISO 8000-1:2022 - Data quality — Part 1: Overview. URL: https://www.iso.org/standard/81745.html.

[21] ISO/IEC 27001:2022 - Information security management systems. URL: https://www.iso.org/standard/27001.

[22] Metadata in Credit Reports: The Key to Smarter More Efficient Reports. URL: https://www.cedar-rose.com/blog/metadata-in-credit-reports-the-key-to-faster-smarter-and-more-efficient-repor

[23] Privacy Statement - Rabobank. URL: https://www.rabobank.nl/en/privacy/privacy-statement.

[24] Regulation - 2016/679 - EN - gdpr - EUR-Lex. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.

[25] Regulation - 2020/1503 - EN - EUR-Lex. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32020R1503.

[26] Regulation - EU - 2023/2859 - EN - EUR-Lex. URL: https://eur-lex.europa.eu/eli/reg/2023/2859/oj/eng.

[27] SCO95 - Glossary and abbreviations. URL: https://www.bis.org/basel_framework/chapter/SCO/95.htm.

[28] Standards | SDMX – Statistical Data and Metadata eXchange. URL: https://sdmx.org/standards-2/.

[29] Terms of Use | Bondora. URL: https://www.bondora.nl/en/terms-of-use/#conflictPolicy.

[30] Unified governance for data and AI | Collibra. URL: https://www.collibra.com/.

[31] What Is Peer-to-Peer (P2P) Lending? Definition and How It Works. URL: https://www.investopedia.com/terms/p/peer-to-peer-lending.asp.

[32] Deutsche Bank Data protection information under the EU General Data Protection Regulation in Germany. 2018.

[33] The ECB Filing and Retention Plan. 2022. URL: https://www.ecb.europa.eu/ecb/access_to_documents/shared/pdf/filing_retention_plan.en.pdf.

[34] Proposal for a Regulation of the European Parliament and of the Council on a framework for Financial Data Access and amending Regulations (EU) No 1093/2010, (EU) No 1094/2010 and (EU) No 1095/2010, COM(2023) 360 final, 6 2023. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52023PC0360.

[35] Final Report on draft Implementing Technical Standards on certain tasks of collection bodies and certain functionalities of the European Single Access Point (ESAP), 10 2024. URL: https://www.eiopa.europa.eu/publications/final-report-draft-implementing-technical-standards-certain-tasks-collection-b en.

[36] Wilhelmina Afua Addy, Chinonye Esther Ugochukwu, Adedoyin Tolulope Oyewole, Onyeka Chrisanctus Ofodile, Omotayo Bukola Adeoye, Chinwe Chinazo Okoye, Wilhelmina Afua Addy, Chinonye Esther Ugochukwu, Adedoyin Tolulope Oyewole, Onyeka Chrisanctus Ofodile, Omotayo Bukola Adeoye, and Chinwe Chinazo Okoye. Predictive analytics in credit risk management for banks: A comprehensive review. *GSC Advanced Research and Reviews*, 18(2):434–449, 2 2024. URL: https://gsconlinepress.com/journals/gscarr/content/predictive-analytics-credit-risk-management-banks-comprehensive-reviewhttps://gsconlinepress.com/journals/gscarr/sites/default/files/GSCARR-2024-0077.pdf, doi:10.30574/GSCARR.2024.18.2.0077.

[37] Habeeb Agoro and P Peace. Assessing the Role of Metadata in Data Governance Policies. 8 2024. URL: https://www.ijsr.net/getabstract.php?paperid=SR220811091304, doi:10.21275/SR220811091304.

[38] Edith Ebele Agu, Njideka Rita Chiekezie, Angela Omozele Abhulimen, and Anwuli Nkemchor Obiki-Osafiele. Harnessing digital transformation to solve operational bottlenecks in banking. *World Journal of Advanced Science and Technology*, 6(1):046–056, 8 2024. URL: https://zealjournals.com/wjast/content/harnessing-digital-transformation-solve-operational-bottlenecks-banking, doi:10.53346/WJAST.2024.6.1.0046.

[39] Ujué Agudo, Karlos G. Liberal, Miren Arrese, and Helena Matute. The impact of AI errors in a human-in-the-loop process. *Cognitive Research: Principles and Implications*, 9(1), 12 2024. URL: https://www.alation.com, doi:10.1186/S41235-023-00529-3.

[40] Mohammad Nadeem Ahmed, Abhijeet Anand, Mohammad Rashid Hussain, Mohammed Mohsin Ahmed, Imran Mohd Khan, and Mohammad Ashiquee Rasool. Artificial Intelligence in Fintech: Emerging Trends and Use Cases. *7th IEEE International Conference on Advanced Technologies, Signal and Image Processing*, 1:459–464, 2024. doi:10.1109/ATSIP62566.2024.10638924.

[41] Rob Aitken. 'All data is credit data': Constituting the unbanked. *Competition and Change*, 21(4):274–300, 8 2017. URL: https://journals-sagepub-com.ezproxy2.utwente.nl/doi/10.1177/1024529417712830, doi:10.1177/1024529417712830/ASSET/6627350B-52CA-4767-9808-B3CFA125B0CE/ASSETS/IMAGES/LARGE/10.1177{_}1024529417712830-FIG3.JPG.

[42] Daniel Gelaw Alemneh and Samantha Kelly Hastings. Exploration of adoption of preservation metadata in cultural heritage institutions: Case of PREMIS. *Proceedings of the ASIST Annual Meeting*, 47(1):1–8, 11 2010. URL: /doi/pdf/10.1002/meet.14504701187https://onlinelibrary.wiley.com/doi/abs/10.1002/meet.14504701187https://asistdl.onlinelibrary.wiley.com/doi/10.1002/meet.14504701187, doi:10.1002/MEET.14504701187;PAGE:STRING:ARTICLE/CHAPTER.

[43] Ebtisam Alharbi, Yojana Gadiya, David Henderson, Andrea Zaliani, Alejandra Delfin-Rossaro, Anne Cambon-Thomsen, Manfred Kohler, Gesa Witt, Danielle Welter, Nick Juty, Caroline Jay, Ola Engkvist, Carole Goble, Dorothy S. Reilly, Venkata Satagopam, Vassilios Ioannidis, Wei Gu, and Philip Gribbon. Selection of data sets for FAIRification in drug discovery and development: Which, why, and how? *Drug Discovery Today*, 27(8):2080–2085, 8 2022. doi:10.1016/J.DRUDIS.2022.05.010.

[44] Ebtisam Alharbi, Rigina Skeva, Nick Juty, Caroline Jay, and Carole Goble. Exploring the Current Practices, Costs and Benefits of FAIR Implementation in Pharmaceutical Research and Development: A Qualitative Interview Study. *Data Intelligence*, 3(4):507–527, 10 2021. URL: https://dx.doi.org/10.1162/dint_a_00109, doi:10.1162/DINT{_}A{_}00109.

[45] Ebtisam Alharbi, Rigina Skeva, Nick Juty, Caroline Jay, and Carole Goble. A FAIR-Decide framework for pharmaceutical R&D: FAIR data cost–benefit assessment. *Drug Discovery Today*, 28(4):103510, 4 2023. doi:10.1016/J.DRUDIS.2023.103510.

[46] Joao Rafael Almeida, Jorge Miguel Silva, and Jose Luis Oliveira. A FAIR Approach to Real-World Health Data Management and Analysis. *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, pages 892–897, 2023. doi:10.1109/CBMS58004.2023.00338.

[47] Celia Alvarez-Romero, Máximo Bernabeu-Wittel, Carlos Luis Parra-Calderón, Silvia Rodríguez Mejías, and Alicia Martínez-García. Desiderata for discoverability and FAIR adoption of health data hubs. *Journal of Biomedical Informatics*, 157:104700, 9 2024. doi:10.1016/J.JBI.2024.104700.

[48] Abn AMRO Bank. Abbreviations and definitions of important terms 2024.

[49] Kushal Ajaybhai Anjaria. Computational implementation and formalism of FAIR data stewardship principles. *Data Technologies and Applications*, 54(2):193–214, 6 2020. doi:10.1108/DTA-09-2019-0164/FULL/XML.

[50] Joan Bader, Chris Hayward, Jonathan Razo, Stuart Madnick, and Michael Siegel. An analysis of data standardization across a capital markets/financial services firm, 1999. URL: https://www.researchgate.net/publication/37594122_An_analysis_of_data_standardization_across_a_capital_marketsfinancial_services_firm.

[51] Bank for International Settlements. Basel Framework. URL: https://www.bis.org/basel_framework/.

[52] Bank for International Settlements. History of the Basel Committee. URL: https://www.bis.org/bcbs/history.htm.

[53] Bank for International Settlements. Principles for effective risk data aggregation and risk reporting. 1 2013. URL: https://www.bis.org/publ/bcbs239.htm.

[54] Tianzi Bao, Yi Ding, Ram Gopal, and Mareike Möhlmann. Throwing Good Money After Bad: Risk Mitigation Strategies in the P2P Lending Platforms. *Information Systems Frontiers*, 26(4):1453–1473, 8 2024. URL: https://link-springer-com.ezproxy2.utwente.nl/article/10.1007/s10796-023-10423-4, doi:10.1007/S10796-023-10423-4/TABLES/4.

[55] Frédérique Belliard, Angelica Maria Maineri, Esther Plomp, Andrés Felipe Ramos Padilla, Junzi Sun, and Maryam Zare Jeddi. Ten simple rules for starting FAIR discussions in your community. *PLOS Computational Biology*, 19(12):1–16, 12 2023. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1011668, doi:10.1371/JOURNAL.PCBI.1011668.

[56] Stefan Bender, Brigitte Hausstein, Christian Hirsch, and Deutsche Bundesbank. An introduction to INEXDA's metadata schema. 2018.

[57] Mike Bennett. The financial industry business ontology: Best practice for big data. *Journal of Banking Regulation*, 14(3-4):255–268, 7 2013. URL: https://link-springer-com.ezproxy2.utwente.nl/article/10.1057/jbr.2013.13, doi:10.1057/JBR.2013.13/METRICS.

[58] Elisa Bertino, Chenyun Dai, and Murat Kantarcioglu. The Challenge of Assuring Data Trustworthiness. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5463:22–33, 2009. URL: https://link.springer.com/chapter/10.1007/978-3-642-00887-0_2, doi:10.1007/978-3-642-00887-0{_}2.

[59] Wentai Bi and Yuan Liang. Risk Assessment of Operator's Big Data Internet of Things Credit Financial Management Based on Machine Learning. *Mobile Information Systems*, 2022(1), 1 2022. URL: https://onlinelibrary.wiley.com/doi/10.1155/2022/5346995, doi:10.1155/2022/5346995.

[60] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked Data - The Story So Far. *Linking the World's Information*, pages 115–143, 7 2023. URL: /doi/pdf/10.1145/3591366.3591378?download=true, doi:10.1145/3591366.3591378.

[61] Gero Friedrich Bone-Winkel and Felix Reichenbach. Improving credit risk assessment in P2P lending with explainable machine learning survival analysis. *Digital Finance*, 6(3):501–542, 9 2024. URL: https://link.springer.com/article/10.1007/s42521-024-00114-3, doi:10.1007/S42521-024-00114-3/FIGURES/14.

[62] Anca Raluca Breje, Robert Gyorödi, Cornelia Gyorödi, Doina Zmaranda, and George Pecherle. Comparative Study of Data Sending Methods for XML and JSON Models. *International Journal of Advanced Computer Science and Applications*, 9(12):198–204, 33 2018. URL: www.ijacsa.thesai.org, doi:10.14569/IJACSA.2018.091229.

[63] Peter Buneman, Sanjeev Khanna, and Wang Chiew Tan. Why and Where: A Characterization of Data Provenance. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1973:316–330, 2001. URL: https://link-springer-com.ezproxy2.utwente.nl/chapter/10.1007/3-540-44503-X_20, doi:10.1007/3-540-44503-X{_}20.

[64] Qiang Cai and Qian Qian. Summary of Credit Risk Assessment Methods. pages 89–93, 8 2018. URL: https://www.atlantis-press.com/proceedings/icitme-18/25906421, doi:10.2991/ICITME-18.2018.18.

[65] Douglas Campbell. Identifying the Identifiers. *Proceedings of the International Conference on Dublin Core and Metadata Applications*, 8 2007. URL: https://dcpapers.dublincore.org/article/952108687, doi:10.23106/DCMI.952108687.

[66] Stefano Cappiello. The EBA and the Banking Union. *European Business Organization Law Review*, 16(3):421–437, 9 2015. URL: https://link.springer.com/article/10.1007/s40804-015-0016-9, doi:10.1007/S40804-015-0016-9/METRICS.

[67] Rebecca Chan, Indrit Troshani, Sally Rao Hill, and Arvid Hoffmann. Towards an understanding of consumers' FinTech adoption: the case of Open Banking. *International Journal of Bank Marketing*, 40(4):886–917, 4 2022. doi:10.1108/IJBM-08-2021-0397/FULL/PDF.

[68] Jiajia Chen and Weiqing Cheng. Analysis of web traffic based on HTTP protocol. *2016 24th International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2016*, 12 2016. doi:10.1109/SOFTCOM.2016.7772120.

[69] Galyna Chornous, Galyna Chornous, and Ganna Ursulenko. RISK MANAGEMENT IN BANKS: NEW APPROACHES TO RISK ASSESSMENT AND INFORMATION SUPPORT. *Ekonomika*, 92(1):120–132, 1 2013. URL: https://www.journals.vu.lt/ekonomika/article/view/1131, doi:10.15388/Ekon.2013.0.1131.

[70] João Constantino, Henrique São Mamede, and Miguel Mira da Silva. Adopting ISO 20022: opportunities, challenges, and success factors for corporations in payment processing. *Emerging Science Journal*, 8(4):1402–1419, 8 2024. URL: http://hdl.handle.net/10400.2/19482https://creativecommons.org/licenses/by/4.0/, doi:10.28991/ESJ-2024-08-04-010.

[71] John W. Creswell and Vicki L. P. Clark. Designing and Conducting Mixed Methods Research. *Sage Publications*, 2017.

[72] Nicholas Dacre, Odai AlJaloudi, M.K.S. Al-Mhdawi, Mouhamed Thiam, Muath Abdel Qader, and Abroon Qazi. Examining the Impact of Credit Risk Management Challenges on Bank Performance: The Mediating Role of Best Practices. *SSRN Electronic Journal*, 11 2024. URL: https://papers.ssrn.com/abstract=5024281, doi:10.2139/SSRN.5024281.

[73] Yugank David. Legal and Regulatory Implications, Challenges and Opportunities in Peer-to-peer Lending Platforms: a Critical and Comparative Analysis. 6 2024. URL: https://papers.ssrn.com/abstract=4922308, doi:10.2139/SSRN.4922308.

[74] Rowdy de Groot, Frank van der Graaff, Daniël van der Doelen, Michiel Luijten, Ronald De Meyer, Hekmat Alrouh, Hedy van Oers, Jacintha Tieskens, Josjan Zijlmans, Meike Bartels, Arne Popma, Nicolette de Keizer, Ronald Cornet, and Tinca J C Polderman. Implementing Findable, Accessible, Interoperable, Reusable (FAIR) Principles in Child and Adolescent Mental Health Research: Mixed Methods Approach. *JMIR mental health*, 11(1), 12 2024. URL: http://www.ncbi.nlm.nih.gov/pubmed/39727091, doi:10.2196/59113.

[75] Natalia Queiroz de Oliveira, Vânia Borges, Henrique F. Rodrigues, Maria Luiza Machado Campos, and Giseli Rabello Lopes. A Practical Approach of Actions for FAIRification Workflows. *Communications in Computer and Information Science*, 1537 CCIS:94–105, 2022. URL: https://link.springer.com/chapter/10.1007/978-3-030-98876-0_8, doi:10.1007/978-3-030-98876-0{_}8/FIGURES/9.

[76] Dize Dinçkol, Pinar Ozcan, and Markos Zachariadis. Regulatory standards and consequences for industry architecture: The case of UK Open Banking. *Research Policy*, 52(6):104760, 7 2023. URL: https://www.sciencedirect.

com/science/article/pii/S0048733323000446, doi:10.1016/J.RESPOL.2023.104760.

[77] David Donoho. 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, 26(4):745–766, 10 2017. URL: https://www.tandfonline.com/doi/pdf/10.1080/10618600.2017.1384734, doi:10.1080/10618600.2017.1384734;CTYPE:STRING:JOURNAL.

[78] Jie Du and Lina Zhou. Improving financial data quality using ontologies. *Decision Support Systems*, 54(1):76–86, 12 2012. URL: https://www.sciencedirect.com/science/article/pii/S0167923612001832?casa_token=RJOYl_6D3fkAAAAA:Drl1xb4JJKl12FXwuup_k8WxsDvBLgLgzObKDzsVB2qc7pquqca36hrptnwyNzo_uW4-LXJe1Q, doi:10.1016/J.DSS.2012.04.016.

[79] Mona A. ElBannan. The Financial Crisis, Basel Accords and Bank Regulations: A Conceptual Framework. *International Journal of Accounting and Financial Reporting*, 7(2):225–275, 12 2017. URL: https://www.macrothink.org/journal/index.php/ijafr/article/view/12122, doi:10.5296/IJAFR.V7I2.12122.

[80] ESMA. Study on data formats and transmission protocols, 1 2024. URL: https://www.esma.europa.eu/document/study-data-formats-and-transmission-protocols.

[81] European Banking Authority. Guidelines on loan origination and monitoring | European Banking Authority, 5 2020. URL: https://www.eba.europa.eu/activities/single-rulebook/regulatory-activities/credit-risk/guidelines-loan-origination-and-monitoring.

[82] European Central Bank. Guideline (EU) 2017/2335 of the European Central Bank of 23 November 2017 on the procedures for the collection of granular credit and credit risk data (ECB/2017/38), 11 2017. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32017O0038.

[83] European Parliament & Council of the European Union. Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC, 11 2015. URL: https://eur-lex.europa.eu/eli/dir/2015/2366/oj/eng.

[84] FAIR Data Maturity Model Working Group. FAIR Data Maturity Model. Specification and Guidelines. 6 2020. URL: https://zenodo.org/records/3909563, doi:10.15497/RDA00050.

[85] Adrienne Porter Felt, Richard Barnes, April King, Chris Palmer, Chris Bentzel, and Parisa Tabriz. Measuring HTTPS Adoption on the Web. *Proceedings of the 26th USENIX Security Symposium*, pages 1323–1338, 2017. URL: http://ipads.se.sjtu.edu.cn/lib/exe/fetch.php?

media=publications:vtz.pdf%0Ahttps://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/hua.

[86] Federico Ferretti. Peer-to-Peer Lending and EU Credit Laws: A Creditworthiness Assessment, Credit-Risk Analysis or … Neither of the Two? *German Law Journal*, 22(1):102–121, 1 2021. URL: https://www.cambridge.org/core/journals/german-law-journal/article/peertopeer-lending-and-eu-credit-laws-a-creditworthiness-assessment-creditrisk
5CE96E68947E0A9496E018D67FD5DD50, doi:10.1017/GLJ.2020.100.

[87] Daniel Fett, Ralf Küsters, and Guido Schmitz. A comprehensive formal security analysis of OAuth 2.0. *Proceedings of the ACM Conference on Computer and Communications Security*, 24-28-October-2016:1204–1215, 10 2016. URL: https://dl.acm.org/doi/pdf/10.1145/2976749.2978385, doi:10.1145/2976749.2978385;SUBPAGE:STRING:ABSTRACT;CSUBTYPE:STRING:CONFERENCE.

[88] Michal S. Gal and Daniel L. Rubinfeld. Data Standardization. *New York University Law Review*, 94, 2019. URL: https://heinonline.org/HOL/Page?handle=hein.journals/nylr94&id=755&div=&collection=.

[89] Richard Gartner. What Metadata Is and Why It Matters. *Metadata*, pages 1–13, 2016. URL: https://link.springer.com/chapter/10.1007/978-3-319-40893-4_1, doi:10.1007/978-3-319-40893-4{_}1.

[90] Alexander L. George and Andrew Bennett. Case studies and theory development in the social sciences. *Belfer Studies in International Security*, 2005.

[91] Boris Glavic and Klaus Dittrich. Data Provenance: A Categorization of Existing Approaches. *GI-Edition - Lecture notes in informatics (LNI). Proceedings*, (103):227–241, 2007. URL: https://dl.gi.de/handle/20.500.12116/31801.

[92] Rocio Gonzalez Soltero, Debora Pino García, Alberto Bellido, Pablo Ryan, and Ana I. Rodríguez-Learte. FAIR data management: a framework for fostering data literacy in biomedical sciences education. *BMC Medical Research Methodology*, 24(1):1–9, 12 2024. URL: https://bmcmedresmethodol-biomedcentral-com.ezproxy2.utwente.nl/articles/10.1186/s12874-024-02404-1, doi:10.1186/S12874-024-02404-1/TABLES/4.

[93] Pankaj Gupta and Sivakumar Ponnusamy. Beyond Banking: The Trailblazing Impact of Data Lakes on Financial Landscape. *International Journal of Computer Applications*, 185(47):24–29, 12 2023. doi:10.5120/IJCA2023923287.

[94] Alex R Hardisty, Wouter Addink, Falko Glöckler, Anton Güntsch, Sharif Islam, and Claus Weiland. A choice of persistent identifier schemes for the Distributed System of Scientific Collections (DiSSCo). *Research Ideas and Outcomes 7: e67379*, 7:e67379–, 7 2021. URL: https://riojournal.com/article/67379/, doi:10.3897/RIO.7.E67379.

[95] Hasan Emre Hayretci and Fatma Başak Aydemir. A Multi Case Study on Legacy System Migration in the Banking Industry. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12751 LNCS:536–550, 2021. URL: https://link.springer.com/chapter/10.1007/978-3-030-79382-1_32, doi:10.1007/978-3-030-79382-1{_}32/TABLES/3.

[96] Randy Heiland, Scott Koranda, Suresh Marru, Marlon Pierce, and Von Welch. Authentication and authorization considerations for a multi-tenant service. *SCREAM 2015 - Proceedings of the 2015 Workshop on the Science of Cyberinfrastructure: Research, Experience, Applications and Models, Part of HPDC 2015*, pages 29–35, 6 2015. URL: https://dl.acm.org/doi/pdf/10.1145/2753524.2753534, doi:10.1145/2753524.2753534;PAGE:STRING:ARTICLE/CHAPTER.

[97] George Hripcsak, Jon D. Duke, Nigam H. Shah, Christian G. Reich, Vojtech Huser, Martijn J. Schuemie, Marc A. Suchard, Rae Woong Park, Ian Chi Kei Wong, Peter R. Rijnbeek, Johan Van Der Lei, Nicole Pratt, G. Niklas Norén, Yu Chuan Li, Paul E. Stang, David Madigan, and Patrick B. Ryan. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Studies in Health Technology and Informatics*, 216:574–578, 2015. URL: https://ebooks.iospress.nl/doi/10.3233/978-1-61499-564-7-574, doi:10.3233/978-1-61499-564-7-574.

[98] Laura D. Hughes, Ginger Tsueng, Jack DiGiovanna, Thomas D. Horvath, Luke V. Rasmussen, Tor C. Savidge, Thomas Stoeger, Serdar Turkarslan, Qinglong Wu, Chunlei Wu, Andrew I. Su, and Lars Pache. Addressing barriers in FAIR data practices for biomedical data. *Scientific Data 2023*, 98(10):1–7, 2 2023. URL: https://www.nature.com/articles/s41597-023-01969-8, doi:10.1038/s41597-023-01969-8.

[99] International Monetary Fund. European Union: Publication of Financial Sector Assessment Program Documentation - Technical Note on European Banking Authority. *IMF Staff Country Reports*, 13(74):1, 2013. doi:10.5089/9781475533781.002.

[100] Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtot, Mercè Crosas, Michel Dumontier, Chris T. Evelo, Carole Goble, Giancarlo Guizzardi, Karsten Kryger Hansen, Ali Hasnain, Kristina Hettne, Jaap Heringa, Rob W.W. Hooft, Melanie Imming, Keith G. Jeffery, Rajaram Kaliyaperumal, Martijn G. Kersloot, Christine R. Kirkpatrick, Tobias Kuhn, Ignasi Labastida, Barbara Magagna, Peter McQuilton, Natalie Meyers, Annalisa Montesanti, Mirjam van Reisen, Philippe Rocca-Serra, Robert Pergl, Susanna Assunta Sansone, Luiz Olavo Bonino da Silva Santos, Juliane Schneider, George Strawn, Mark Thompson, Andra Waagmeester, Tobias Weigel, Mark D. Wilkinson, Egon L. Willighagen, Peter Wittenburg, Marco Roos, Barend Mons, and Erik Schultes.

FAIR Principles: Interpretations and Implementation Considerations. *Data Intelligence*, 2(1-2):10–29, 1 2020. URL: https://dx-doi-org.ezproxy2.utwente.nl/10.1162/dint_r_00024, doi:10.1162/DINT{_}R{_}00024.

[101] Annika Jacobsen, Rajaram Kaliyaperumal, Luiz Olavo Bonino da Silva Santos, Barend Mons, Erik Schultes, Marco Roos, and Mark Thompson. A Generic Workflow for the Data FAIRification Process. *Data Intelligence*, 2(1-2):56–65, 1 2020. URL: https://dx.doi.org/10.1162/dint_a_00028, doi:10.1162/DINT{_}A{_}00028.

[102] Ravi Jagadish. Identity and Access Management in the Cloud: Strengthening Authentication and Authorization for Financial Tech, 2024. URL: https://ijctjournal.org/identity-access-management-cloud-financial-tech/.

[103] Julapa Jagtiani and Catharine Lemieux. The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform. *Financial Management*, 48(4):1009–1029, 12 2019. URL: https://onlinelibrary-wiley-com.ezproxy2.utwente.nl/doi/full/10.1111/fima.12295https://onlinelibrary-wiley-com.ezproxy2.utwente.nl/doi/abs/10.1111/fima.12295https://onlinelibrary-wiley-com.ezproxy2.utwente.nl/doi/10.1111/fima.12295, doi:10.1111/FIMA.12295.

[104] Tanja Jørgensen. Peer-to-Peer Lending – A New Digital Intermediary, New Legal Challenges. *Nordic Journal of Commercial Law*, (1):30–30, 11 2018. URL: https://journals.aau.dk/index.php/NJCL/article/view/2491, doi:10.5278/OJS.NJCL.V0I1.2491.

[105] Anna Jurek and P. Deepak. It pays to be certain: Unsupervised record linkage via ambiguity minimization. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10939 LNAI:177–190, 2018. URL: https://link.springer.com/chapter/10.1007/978-3-319-93040-4_15, doi:10.1007/978-3-319-93040-4{_}15/TABLES/4.

[106] Evangelos Kalapodas and Mary E. Thomson. Credit risk assessment: a challenge for financial institutions. *IMA Journal of Management Mathematics*, 17(1):25–46, 1 2006. URL: https://dx.doi.org/10.1093/imaman/dpi026, doi:10.1093/IMAMAN/DPI026.

[107] Man Ying Kang and Marcel Ausloos. An Inverse Problem Study: Credit Risk Ratings as a Determinant of Corporate Governance and Capital Structure in Emerging Markets: Evidence from Chinese Listed Companies. *Economies*, 5(4):47, 11 2017. URL: https://www.mdpi.com/2227-7099/5/4/47/htmhttps://www.mdpi.com/2227-7099/5/4/47, doi:10.3390/ECONOMIES5040047.

[108] Amin Karami and Chukwuemeka Igbokwe. The impact of big data characteristics on credit risk assessment. *International Journal of Data Science and Analytics 2025*, pages 1–21, 3 2025. URL: https://link.springer.com/article/10.1007/s41060-025-00753-8, doi:10.1007/S41060-025-00753-8.

[109] Amandeep Kaur, Srinidhi Ayyagari, Manasi Mishra, and Rachit Thukral. A Literature Review on Device-to-Device Data Exchange Formats for IoT Applications. *JOURNAL OF INTELLIGENT SYSTEMS AND COMPUTING*, 1(1):1–10, 12 2020. URL: https://scienceandtech.co.uk/journals/index.php/jiscom/article/view/4.

[110] Vineela Komandla. Navigating Open Banking: Strategic Impacts on Fintech Innovation and Collaboration. *International Journal of Science and Research (IJSR)*, 6(9):2049–2056, 9 2017. URL: https://papers.ssrn.com/abstract=4982992, doi:10.21275/SR24913161731.

[111] Hariharan Pappil Kothandapani. Optimizing Financial Data Governance for Improved Risk Management and Regulatory Reporting in Data Lakes. *International Journal of Applied Machine Learning and Computational Intelligence*, 1 2022. doi:http://dx.doi.org/10.6084/m9.figshare.28234490.v1.

[112] Hemanth Kumar and Msc In Cs. Leveraging Cloud Computing for Scalable Financial Technology Solutions. *IJSAT - International Journal on Science and Technology*, 11(1), 2020. URL: https://www.ijsat.org/research-paper.php?id=1570, doi:10.5281/ZENODO.14752357.

[113] Manish Kumar. Data governance journey at SEA's largest SME digital financing platform, 2021. URL: https://medium.com/fsmk-engineering/data-governance-journey-at-seas-largest-digital-p2p-lending-platform-part-2-e

[114] Wahyu Haris Kusuma Atmaja, Mohammad Isa, and Muhammad Aidil Fahmy. Household Electricity Profile as Alternative Data for Credit Appraisal. *ACM International Conference Proceeding Series*, pages 350–355, 11 2022. URL: https://dl-acm-org.ezproxy2.utwente.nl/doi/10.1145/3575882.3575950, doi:10.1145/3575882.3575950.

[115] Clement Labadie, Christine Legner, Markus Eurich, and Martin Fadler. FAIR Enough? Enhancing the Usage of Enterprise Data with Data Catalogs. *Proceedings - 2020 IEEE 22nd Conference on Business Informatics, CBI 2020*, 1:201–210, 6 2020. doi:10.1109/CBI49978.2020.00029.

[116] Ignasi Labastida and Thomas Margoni. Licensing FAIR Data for Reuse. *Data Intelligence*, 2(1-2):199–207, 1 2020. URL: https://dx-doi-org.ezproxy2.utwente.nl/10.1162/dint_a_00042, doi:10.1162/DINT{_}A{_}00042.

[117] Annalisa Landi, Mark Thompson, Viviana Giannuzzi, Fedele Bonifazi, Ignasi Labastida, Luiz Olavo Bonino da Silva Santos, and Marco Roos. The "A" of FAIR – As Open as Possible, as Closed as Necessary. *Data Intelligence*,

2(1-2):47–55, 1 2020. URL: https://dx-doi-org.ezproxy2.utwente.nl/10.1162/dint_a_00027, doi:10.1162/DINT{_}A{_}00027.

[118] Xinlong Li and Andrew T. Ching. How Does a Firm Adapt in a Changing World? The Case of Prosper Marketplace. *Marketing Science*, 43(3):673–693, 5 2024. URL: /doi/pdf/10.1287/mksc.2022.0198, doi:10.1287/MKSC.2022.0198;WEBSITE:WEBSITE:INFORMS-SITE.

[119] Zihuai Ma. The Investigation of Communications Protocol. *Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*, pages 576–582, 2 2024. URL: https://www.atlantis-press.com/proceedings/dai-23/125998083, doi:10.2991/978-94-6463-370-2{_}58.

[120] Syed Rahman Mashwani, Azhar Rauf, Shah Khusro, and Saeed Mahfooz. Linked file system: Towards exploiting linked data technology in file systems. *ICOSST 2016 - 2016 International Conference on Open Source Systems and Technologies, Proceedings*, pages 135–141, 1 2017. doi:10.1109/ICOSST.2016.7838590.

[121] Mykola Melnyk, Mykyta Kuchkin, and Andriy Blyznyukov. Commercial Banks: Traditional Banking Models Vs. Fintechs Solutions. *Financial Markets, Institutions and Risks*, 6(2):122–129, 2022. URL: https://essuir.sumdu.edu.ua/handle/123456789/88535, doi:10.21272/FMIR.6(2).122-129.2022.

[122] Haris Memi´c Memi´c, Alen Bošnjakovi´cbošnjakovi´c, Džemi´c Džemi´c, Almir Badnjevi´c Badnjevi´c, H Memi´cmemi´c, · A Bošnjakovi´cbošnjakovi´c, · Z Džemi´cdžemi´c, and A Badnjevi´c Badnjevi´c. Regulations and Directives—Past, Present, Future. pages 5–31, 2024. URL: https://link.springer.com/chapter/10.1007/978-3-031-43444-0_2, doi:10.1007/978-3-031-43444-0{_}2.

[123] Helen Tadesse Moges, Karel Dejaeger, Wilfried Lemahieu, and Bart Baesens. A multidimensional analysis of data quality for credit risk management: New insights and challenges. *Information & Management*, 50(1):43–58, 1 2013. doi:10.1016/J.IM.2012.10.001.

[124] Barend Mons, Cameron Neylon, Jan Velterop, Michel Dumontier, Luiz Olavo Bonino Da Silva Santos, and Mark D. Wilkinson. Cloudy, increasingly FAIR; Revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services and Use*, 37(1):49–56, 2 2017. URL: https://journals.sagepub.com/doi/full/10.3233/ISU-170824, doi:10.3233/ISU-170824/ASSET/IMAGES/LARGE/10.3233{_}ISU-170824-FIG1.JPEG.

[125] Barend Mons, Erik Schultes, Fenghong Liu, and Annika Jacobsen. The FAIR Principles: First Generation Implementation Choices and Challenges. *Data Intelligence*, 2(1-2):1–9, 1 2020. URL: https://dx-doi-org.ezproxy2.utwente.nl/10.1162/dint_e_00023, doi:10.1162/DINT{_}E{_}00023.

[126] Mark A. Musen, Martin J. O'Connor, Erik Schultes, Marcos Martínez-Romero, Josef Hardi, and John Graybeal. Modeling community standards for metadata as templates makes data FAIR. *Scientific Data 2022 9:1*, 9(1):1–15, 11 2022. URL: https://www.nature.com/articles/s41597-022-01815-3, doi:10.1038/s41597-022-01815-3.

[127] Janifer Nahar, Md Atiqur Rahaman, Md Alauddin, Farhana Zaman Rozony, and Predictive Analytics. Big Data in Credit Risk Management: a Systematic Review of Transformative Practices and Future Directions. *International Journal of Management Information Systems and Data Science*, 1(04):68–79, 9 2024. URL: https://globalmainstreamjournal.com/index.php/IJMISDS/article/view/196, doi:10.62304/IJMISDS.V1I04.196.

[128] Andy Neumann, Nuno Laranjeiro, and Jorge Bernardino. An Analysis of Public REST Web Service APIs. *IEEE Transactions on Services Computing*, 14(4):957–970, 7 2021. doi:10.1109/TSC.2018.2847344.

[129] Mikael Nilsson and Qionglin Shan. Credit risk analysis with machine learning techniques in peer-to-peer lending market. 2018. URL: https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-176450.

[130] Nurzhan Nurseitov, Michael Paulson, Randall Reynolds, and C. Izurieta. Comparison of JSON and XML Data Interchange Formats: A Case Study. *ISCA International Conference on Computer Applications in Industry and Engineering*, 2009.

[131] Adedotun Oladinni and Goodness Tolulope Adewale. Innovative API Frameworks and Data-Driven Modelling for Enhanced FinTech Lending Applications. *International Research Journal of Modernization in Engineering Technology and Science*, 7(1), 1 2025. URL: https://www.researchgate.net/publication/387872164_Innovative_API_Frameworks_and_Data-Driven_Modeling_for_Enhanced_FinTech_Lending_Applications, doi:10.56726/IRJMETS66012.

[132] Bruno Oliveira, Ana Duarte, and Óscar Oliveira. Towards a Data Catalog for Data Analytics. *Procedia Computer Science*, 237:691–700, 1 2024. URL: https://www.sciencedirect.com/science/article/pii/S1877050924011712, doi:10.1016/J.PROCS.2024.05.155.

[133] Anna Eugenia Omarini. Banks and fintechs: how to develop a digital open banking approach for the bank's future. *INTERNATIONAL BUSINESS RESEARCH*, 11(9):23–36, 8 2018. URL: https://iris.unibocconi.it/handle/11565/4013970, doi:10.5539/IBR.V11N9P23.

[134] Joerg Orgeldinger. The Implementation of Basel Committee BCBS 239: Short analysis of the new rules for Data Management. *Journal of Central Banking Theory and Practice*, 7(3):57–72, 2018. doi:10.2478/JCBTP-2018-0023.

[135] Simon O'Toole and James Tocknell. FAIR standards for astronomical data. *ADASS*, 3 2022. URL: https://arxiv.org/abs/2203.10710v1.

[136] Irene V. Pasquetto, Zoë Cullen, Andrea Thomer, and Morgan Wofford. What is research data "misuse"? And how can it be prevented or mitigated? *Journal of the Association for Information Science and Technology*, 75(12):1413–1429, 12 2024. URL: /doi/pdf/10.1002/asi.24944https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.24944https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24944, doi:10.1002/ASI.24944;PAGE:STRING:ARTICLE/CHAPTER.

[137] Galena Pisoni, Bálint Molnár, and Ádám Tarcsi. Knowledge Management and Data Analysis Techniques for Data-Driven Financial Companies. *Journal of the Knowledge Economy*, 15(3):13374–13393, 11 2023. URL: https://link.springer.com/article/10.1007/s13132-023-01607-z.

[138] Núria Queralt-Rosinach, Rajaram Kaliyaperumal, César H. Bernabé, Qinqin Long, Simone A. Joosten, Henk Jan van der Wijk, Erik L.A. Flikkenschild, Kees Burger, Annika Jacobsen, Barend Mons, and Marco Roos. Applying the FAIR principles to data in a hospital: challenges and opportunities in a pandemic. *Journal of Biomedical Semantics*, 13(12):1–19, 12 2022. URL: https://jbiomedsem-biomedcentral-com.ezproxy2.utwente.nl/articles/10.1186/s13326-022-00263-7http://creativecommons.org/publicdomain/zero/1.0/, doi:10.1186/S13326-022-00263-7/FIGURES/7.

[139] Linda Rainey, Jennifer E. Lutomski, and Mireille J.M. Broeders. FAIR data sharing: An international perspective on why medical researchers are lagging behind. *Big Data and Society*, 10(1), 1 2023. URL: https://journals.sagepub.com/doi/full/10.1177/20539517231171052, doi:10.1177/20539517231171052/ASSET/IMAGES/LARGE/10.1177{_}20539517231171052-FIG2.JPEG.

[140] Marc Rothemund and Maria Gerhardt. The European credit information landscape. An analysis of a survey of credit bureaus in Europe. ECRI Research Report, January 2011. *Archive of European Integration*, 2011. URL: http://shop.ceps.eu/book/european-credit-information-landscape.

[141] Farhana Zaman Rozony, Mst Nahida Aktar Aktar, Md Ashrafuzzaman, and Ashraful Islam. A Systematic Review Of Big Data Integration Challenges And Solutions For Heterogeneous Data Sources. *Academic Journal on Business Administration, Innovation & Sustainability*, 4(04):1–18, 10 2024. URL: https://allacademicresearch.com/index.php/AJBAIS/article/view/111, doi:10.69593/AJBAIS.V4I04.111.

[142] Keith Russell. FAIR Principles and Why They Matter. *Digital Transformation of the Laboratory: A Practical Guide to the Connected Lab*, pages 101–105, 6 2021. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/9783527825042.ch5https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527825042.ch5https://onlinelibrary.wiley.com/doi/10.1002/9783527825042.ch5, doi:10.1002/9783527825042.CH5.

[143] Ravi S. Sandhu and Pierangela Samarati. Access Control: Principles and Practice. *IEEE Communications Magazine*, 32(9):40–48, 1994. doi:10.1109/35.312842.

[144] Nuno Pessanha Santos. The Expansion of Data Science: Dataset Standardization. *Standards 2023, Vol. 3, Pages 400-410*, 3(4):400–410, 11 2023. URL: https://www.mdpi.com/2305-6703/3/4/28/htmhttps://www.mdpi.com/2305-6703/3/4/28, doi:10.3390/STANDARDS3040028.

[145] Andreas Schaad, Jonathan Moffett, and Jeremy Jacob. The role-based access control system of a European bank. *Proceedings of the sixth ACM symposium on Access control models and technologies*, pages 3–9, 5 2001. URL: https://dl.acm.org/doi/pdf/10.1145/373256.373257, doi:10.1145/373256.373257.

[146] Eva Schlehahn and Rigo Wenning. GDPR transparency requirements and data privacy vocabularies. *IFIP Advances in Information and Communication Technology*, 547:95–113, 2019. URL: https://link-springer-com.ezproxy2.utwente.nl/chapter/10.1007/978-3-030-16744-8_7, doi:10.1007/978-3-030-16744-8{_}7/FIGURES/2.

[147] Rizwana Shaikh and M. Sasikumar. Data Classification for Achieving Security in Cloud Computing. *Procedia Computer Science*, 45(C):493–498, 1 2015. URL: https://www.sciencedirect.com/science/article/pii/S1877050915003233, doi:10.1016/J.PROCS.2015.03.087.

[148] Md Labib Sharar. Online Peer-to-Peer Lending: Determinants of Loan Performance. *SSRN Electronic Journal*, 6 2021. URL: https://papers.ssrn.com/abstract=3785323, doi:10.2139/SSRN.3785323.

[149] Daniel Simeon-Dubach, Zisis Kozlakidis, Juhi Tayal, Shannon J. McCall, Wohaib Hasan, Fay Betsou, Jonathan Lawson, and Dominic Allen. Experts Speak Forum: Implementation of the FAIR Principles in Biobanking Needs Fair Incentives. *Biopreservation And Biobanking*, 22(6):557–562, 12 2024. URL: https://www.liebertpub.com/doi/10.1089/bio.2024.0153, doi:10.1089/BIO.2024.0153.

[150] Yogesh L. Simmhan, Beth Plale, and Dennis Gannon. A survey of data provenance techniques. 2005.

[151] A. Anil Sinaci, Francisco J. Núñez-Benjumea, Mert Gencturk, Malte Levin Jauer, Thomas Deserno, Catherine Chronaki, Giorgio Cangioli, Carlos Cavero-Barca, Juan M. Rodríguez-Pérez, Manuel M. Pérez-Pérez, Gokce B. Laleci Erturkmen, Tony Hernández-Pérez, Eva Méndez-Rodríguez, and Carlos L. Parra-Calderón. From Raw Data to FAIR Data: The FAIRification Workflow for Health Research. *Methods of Information in Medicine*, 59(1):21–32, 6 2020. URL: http://www.thieme-connect.de/products/ejournals/html/10.1055/s-0040-1713684http://www.thieme-connect.de/DOI/DOI?10.1055/s-0040-1713684, doi:10.1055/S-0040-1713684/ID/JR19020007-14/BIB.

[152] Ranjeet Kumar Singh, Akanksha Nagpal, Arun Jadhav, and Devika P. Madalli. Assessment of FAIR (Findability, Accessibility, Interoperability, and Reusability) data implementation frameworks: a parametric approach. 12 2024. URL: https://arxiv.org/pdf/2504.06268.

[153] Marilyn Skrocki. Standardization Needs for Effective Interoperability. *Transactions of the International Conference on Health Information Technology Advancement*, 2(1), 10 2013. URL: https://scholarworks.wmich.edu/ichita_transactions/32.

[154] Satyaveda Somepalli. Breaking Down Silos: The Benefits of Interoperable EHRs. *Journal of Scientific and Engineering Research*, 2021(10):244–249, 2021. URL: www.jsaer.com.

[155] Silvia B. Southwick, Cory K. Lampert, and Richard Southwick. Preparing Controlled Vocabularies for Linked Data: Benefits and Challenges. *Journal of Library Metadata*, 15(3-4):177–190, 10 2015. URL: https://www.tandfonline.com/doi/pdf/10.1080/19386389.2015.1099983, doi:10.1080/19386389.2015.1099983;CSUBTYPE:STRING:SPECIAL;PAGE:STRING:ARTICLE/CHAPTER.

[156] Reni Sulastri and Marijn Janssen. Challenges in designing an inclusive Peer-to-peer (P2P) lending system. *ACM International Conference Proceeding Series*, pages 55–65, 7 2023. URL: https://dl.acm.org/doi/10.1145/3598469.3598475, doi:10.1145/3598469.3598475.

[157] Huan Tang. Peer-to-Peer Lenders Versus Banks: Substitutes or Complements? *The Review of Financial Studies*, 32(5):1900–1938, 5 2019. URL: https://dx.doi.org/10.1093/rfs/hhy137, doi:10.1093/RFS/HHY137.

[158] N. L. Taranath and A. Geetha. Credit Risk Evaluation using Decision Support System. *2nd IEEE International Conference on Advances in Information Technology*, 2024. doi:10.1109/ICAIT61638.2024.10690699.

[159] Aleksandra Tarkowska, Denise Carvalho-Silva, Charles E. Cook, Edd Turner, Robert D. Finn, and Andrew D. Yates. Eleven quick tips to build a usable REST API for life sciences. *PLoS Computational Biology*, 14(12):e1006542, 12 2018. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC6292566/, doi:10.1371/JOURNAL.PCBI.1006542.

[160] Jan Top, Sander Janssen, Hendrik Boogaard, Rob Knapen, and Görkem Şimşek-Şenel. Cultivating FAIR principles for agri-food data. *Computers and Electronics in Agriculture*, 196:106909, 5 2022. doi:10.1016/J.COMPAG.2022.106909.

[161] Alina Trifan and José Luís Oliveira. Towards a More Reproducible Biomedical Research Environment: Endorsement and Adoption of the FAIR Principles. *Communications in Computer and Information Science*, 1211:453–470, 2020. URL: https://link-springer-com.

ezproxy2.utwente.nl/chapter/10.1007/978-3-030-46970-2_22, doi:10.1007/978-3-030-46970-2{_}22/TABLES/3.

[162] Albert I. Ugochukwu and Peter W.B. Phillips. Open data ownership and sharing: Challenges and opportunities for application of FAIR principles and a checklist for data managers. *Journal of Agriculture and Food Research*, 16:101157, 6 2024. doi:10.1016/J.JAFR.2024.101157.

[163] Miklos A. Vasarhelyi, David Y. Chan, and J. P. Krahel. Consequences of XBRL Standardization on Financial Statement Data. *Journal of Information Systems*, 26(1):155–167, 3 2012. URL: https://dx.doi.org/10.2308/isys-10258, doi:10.2308/ISYS-10258.

[164] Eric Viardot, Ian Mccarthy, and Jin Chen. Editorial: Standardization in a Digital and Global World: State-of-the-Art and Future Perspectives. *IEEE Transactions on Engineering Management*, 68(1):11–17, 2 2021. doi:10.1109/TEM.2020.3006302.

[165] Huaiqing Wang, Kun Chen, Wei Zhu, and Zhenxia Song. A process model on P2P lending. *Financial Innovation*, 1(1):1–8, 12 2015. URL: https://jfin-swufe.springeropen.com/articles/10.1186/s40854-015-0002-9, doi:10.1186/S40854-015-0002-9/FIGURES/8.

[166] S. Weibel, J. Kunze, C. Lagoze, and M. Wolf. Dublin Core Metadata for Resource Discovery. 9 1998. URL: /doi/pdf/10.17487/RFC2413, doi:10.17487/RFC2413.

[167] Tobias Weigel, Ulrich Schwardmann, Jens Klump, Sofiane Bendoukha, and Robert Quick. Making Data and Workflows Findable for Machines. *Data Intelligence*, 2(1-2):40–46, 1 2020. URL: https://dx-doi-org.ezproxy2.utwente.nl/10.1162/dint_a_00026, doi:10.1162/DINT{_}A{_}00026.

[168] Danielle Welter, Nick Juty, Philippe Rocca-Serra, Fuqi Xu, David Henderson, Wei Gu, Jolanda Strubel, Robert T. Giessmann, Ibrahim Emam, Yojana Gadiya, Tooba Abbassi-Daloii, Ebtisam Alharbi, Alasdair J.G. Gray, Melanie Courtot, Philip Gribbon, Vassilios Ioannidis, Dorothy S. Reilly, Nick Lynch, Jan Willem Boiten, Venkata Satagopam, Carole Goble, Susanna Assunta Sansone, and Tony Burdett. FAIR in action - a flexible framework to guide FAIRification. *Scientific Data 2023*, 10(1):1–9, 5 2023. URL: https://www.nature.com/articles/s41597-023-02167-2, doi:10.1038/s41597-023-02167-2.

[169] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. t Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson,

Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan Van Der Lei, Erik Van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data 2016 3:1*, 3(160018):2052–4463, 3 2016. URL: https://www.nature.com/articles/sdata201618, doi:10.1038/sdata.2016.18.

[170] Mark D. Wilkinson, Michel Dumontier, Susanna Assunta Sansone, Luiz Olavo Bonino da Silva Santos, Mario Prieto, Dominique Batista, Peter McQuilton, Tobias Kuhn, Philippe Rocca-Serra, Merce E. Crosas, and Erik Schultes. Evaluating FAIR maturity through a scalable, automated, community governed framework. *Scientific Data*, 6(1):1–12, 12 2019. URL: https://www.nature.com/articles/s41597-019-0184-5.

[171] Mark D. Wilkinson, Susanna Assunta Sansone, Erik Schultes, Peter Doorn, Luiz Olavo Bonino Da Silva Santos, and Michel Dumontier. A design framework and exemplar metrics for FAIRness. *Scientific Data 2018 5:1*, 5(1):1–4, 6 2018. URL: https://www.nature.com/articles/sdata2018118, doi:10.1038/sdata.2018.118.

[172] Andrew Wilson. How much is enough: Metadata for preserving digital data. *Journal of Library Metadata*, 10(2-3):205–217, 2010. URL: https://www.tandfonline.com/doi/pdf/10.1080/19386389.2010.506395, doi:10.1080/19386389.2010.506395;REQUESTEDJOURNAL:JOURNAL:WJLM20.

[173] Jennifer J. Xu, Yong Lu, and Michael Chau. P2P Lending Fraud Detection: A Big Data Approach. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9074:71–81, 2015. URL: https://link.springer.com/chapter/10.1007/978-3-319-18455-5_5, doi:10.1007/978-3-319-18455-5{_}5.

[174] Eunjung Yeo and Jooyong Jun. Peer-to-Peer Lending and Bank Risks: A Closer Look. *Sustainability 2020, Vol. 12, Page 6107*, 12(15):6107, 7 2020. URL: https://www.mdpi.com/2071-1050/12/15/6107/htmhttps://www.mdpi.com/2071-1050/12/15/6107, doi:10.3390/SU12156107.

[175] Hongke Zhao, Yong Ge, Qi Liu, Guifeng Wang, Enhong Chen, and Hefu Zhang. P2P Lending Survey: Platforms, Recent Advances and Prospects. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(6):1–28, 7 2017. URL: https://dl-acm-org.ezproxy2.utwente.nl/doi/10.1145/3078848, doi:10.1145/3078848.

[176] Anneke Zuiderwijk, Keith Jeffery, and Marijn Janssen. The Potential of Metadata for Linked Open Data and its Value for Users and Publishers. *JeDEM - eJournal of eDemocracy and Open Government*, 4(2):222–244, 12 2012. URL: https://jedem.org/index.php/jedem/article/view/138, doi:10.29379/jedem.v4i2.138.

Appendix A

FAIR Assessment Traditional Banks Using Open Documentation and Literature

This appendix presents the FAIRness assessment for traditional banks. The evaluation is informed by a combination of literature, regulatory frameworks, institutional guidelines and related documentation. It is assumed that banks operate in alignment with the regulatory standards relevant to the European financial sector and often align with published metadata standards. The assessment was performed using the original version of the FAIR evaluation framework, as described in Chapter 5. Each FAIR sub-principle is evaluated across specific criteria. Each criterion is scored either non-compliant (0) or compliant (1). The average per sub-principle represents the degree of alignment with the FAIR principles.

Table A.1: FAIR Assessment Traditional Banks Using Open Documentation and Literature

ID	Target	Criterion	Score	Justification
Findability				
F1-1	Data	Each data object is assigned an identifier that is unique at the institutional level. Where possible, globally unique identifiers, like the Legal Entity Identifier (LEI), are used.	1	Banks are required by Regulation (EU) 2016/867 (AnaCredit) to assign unique identifiers to data objects and to use globally unique identifiers such as the LEI for counterparties where available [82].
F1-2	Data	Identifiers remain stable and unchanged throughout the data lifecycle, this includes among others system migrations, updates and archiving.	1	Assuming banks comply with the AnaCredit Regulation, data identifiers are kept stable over time and are not changed as required by the regulation [16].
F1-3	Metadata	Each metadata record is assigned an identifier that is unique within the institution.	1	Guideline (EU) 2017/2335 requires that National Central Banks provide AnaCredit data to the ECB in accordance with SDMX standards [82]. The SDMX standard specifies that each metadata record includes a unique identifier [28]. Since banks must generate and maintain metadata conforming to SDMX when reporting to National Central Banks, it is reasonable to conclude that internal metadata structures also implement SDMX and thus institutionally unique identifiers.
F1-4	Metadata	Identifiers remain stable and unchanged throughout the data lifecycle, this includes among others system migrations, updates and archiving.	1	Similar to data identifiers, metadata identifiers must also remain stable over time to meet AnaCredit requirements [16].
Average Score:			**1**	

F2-1	Metadata	Metadata includes core descriptive elements hat support effective search and discovery, such as title, description, keywords, format and classification terms.	1	The International Network for Exchanging Experience on Statistical Handling of Granular Data (INEXDA) is a collaborative initiative among central banks and international financial institutions aimed at harmonizing practices for managing and documenting granular datasets [15, 56]. SDMX is widely used for structured statistical reporting and regulatory submissions, but is primarily focused on aggregated data and provides limited support for extended variable-level descriptions. In contrast, INEXDA defines a metadata schema for documenting more granular financial datasets. INEXDA includes key descriptive elements such as the name of the dataset, descriptions, keywords, unit descriptions, collection mode and keywords [56]. This schema has been adopted by INEXDA member institutions as a common standard for metadata documentation [15, 56]. Under INEXDA's metadata specifications this criterion is met.
F2-2	Metadata	Metadata includes basic provenance elements (e.g., origin, creator, creation date) that support findability.	1	INEXDA defines a metadata schema that includes core provenance elements such as "Creator" and "Publication Date". These elements document who created the dataset, who is responsible and when it was released [15, 56]. Although detailed transformation histories are not explicitly modeled, the schema supports important provenance elements related to findability.
Average Score:			**1**	

F3-1	Metadata	Metadata includes an explicit reference to the corresponding dataset using a unique and stable identifier.	1	As mentioned earlier banks reporting to the ECB are required to provide data in accordance with the SDMX standard [82]. As part of this standard, metadata must include a stable and unique identifier explicitly linking it to the corresponding dataset [28]. Given this requirement, it can reasonably be assumed that banks generate metadata in accordance with this criterion.
F3-2	Metadata	The linkage between metadata and data is represented in a structured, machine-readable format.	1	SDMX ensures metadata is represented in structured, machine-readable formats, such as XML or JSON. Separately, within the INEXDA network, institutions agreed on using the GESIS metadata tool for the collection of metadata [15]. The GESIS metadata tool generates metadata in a structured, machine-readable format, based on the Data Documentation Initiative (DDI) standard [14]. GESIS explicitly states that its metadata practices algin with the FAIR principles [14]. While SDMX ensures compliance with formal reporting, the GESIS approach reflects best practices for documenting internal detailed financial data. Both indicate that structured, machine-readable metadata is supported in the banking domain.
Average Score:			**1**	

F4-1	Data	Data is registered/indexed in a searchable resource.	1	Some major banks are known to use commercial data catalog tools such as Collibra, Alation, and Informatica. Assuming other banks implement similar tools (including custom-built) internally, credit risk data is generally indexed in such catalogs [5, 30, 39].
F4-2	Metadata	Metadata is registered/indexed in a searchable resource.	1	In the data catalog systems like Collibra, Alation and Informatica, metadata is indexed alongside the corresponding datasets. These platforms are designed to make both data and metadata searchable through an interface, ensuring discoverability [5, 30, 39].
F4-3	Infrastructure	The searchable resource supports keyword or attribute-based search across datasets and metadata records.	1	Tools like Collibra and Alation offer structured, attribute-based and free-text search interfaces. This allows for data and metadata search, based on filters [30, 39].
F4-4	Infrastructure	The searchable resource provides comprehensive coverage of all relevant datasets and metadata.	0	Even though catalog tools can support full coverage, in reality not all credit risk datasets are consistently indexed, for example because of legacy systems [95].
F4-5	Infrastructure	The searchable resource provides a machine-readable interface for automated search and discovery.	1	Tools like Collibra, Informatica and Alation provide REST APIs and metadata query services [5, 30, 39]. Given that banks choose these platforms, it is reasonable to assume that such APIs are actively used, making the searchable resource machine-readable.
Average Score:			**0.8**	

Accessibility

A1-1	Data	Data can be retrieved using their identifier via a communication protocol (e.g., HTTPS or FTP).	1	Banks provide access to credit data through standardized methods. Some major Dutch banks, like Rabobank, ABN Amro and ING, like many other financial institutions, use APIs that operate over HTTPS that allow data to be retrieved using identifiers [2, 7, 8].
A1-2	Metadata	Metadata is retrievable using its identifier via the same communication protocol as the associated data.	1	Banks typically provide metadata alongside data through the same access channels. The APIs of the Dutch banks, for example, allow users to retrieve both data and its metadata consistently in the same manner [2, 7, 8].
A1-3	Infrastructure	The communication protocol used to access (meta)data is well-documented and based on established standards.	1	Banks use standardized and well-documented communication protocols for data retrieval. For example, the APIs from the Dutch banks [2, 7, 8], are properly documented and use HTTPS, a globally recognized and standardized protocol.
Average Score:			**1**	

A1.1-1	Infrastructure	Data retrieval relies on open, non-proprietary communication protocols that are freely and universally implementable.	1	Many institution make use of APIs over HTTPs, which is an open and freely implementable protocol not tied to proprietary software [2, 7, 8].
A1.1-2	Infrastructure	The communication protocols used are publicly documented and accessible.	1	The API documentations describing among others endpoints, request/response structures and protocol behavior are publicly documented [2, 7, 8]. These comprehensive documentations support consistent implementation and integration.

A1.1-3	Infrastructure	The same protocol is consistently applied across systems and departments within the institution.	1	Institutions such as Rabobank, ABN Amro, and ING apply HTTPS consistently across their API ecosystems [2, 7, 8].
A1.1-4	Infrastructure	The protocol allows for automated data retrieval without requiring manual intervention.	1	Once authorized, data can be retrieved programmatically, without the need for manual entry [2, 7, 8]. This enables integration into automated workflows.
Average Score:			**1**	

A1.2-1	Infrastructure	The communication protocol supports authentication to verify the identity of users or systems accessing (meta)data.	1	Banks implement strong authentication mechanisms. For example, banks use approaches such as mutual TLS and client credentials to authenticate API clients [2, 7, 8].
A1.2-2	Infrastructure	The protocol supports authorization mechanisms to control access to (meta)data.	1	The APIs used by banks implement access control through OAuth 2.0 [2, 7, 8]. Thereby they restrict access to specific operations based on the user's authorization level.
A1.2-3	Infrastructure	Access control mechanisms comply with relevant legal, regulatory and institutional requirements	1	Banks in the European financial sector are legally required to comply with regulations such as PSD2 and the GDPR. These frameworks mandate strong access control and data protection mechanisms [24, 83]. Banks implement access controls that align with these regulatory obligations, ensuring legal compliance.
Average Score:			**1**	

A2-1	Metadata	Metadata remains accessible through the same infrastructure, even after the associated dataset is archived or deleted.	1	Financial institutions must comply with strict auditability and traceability requirements under frameworks such as Basel III, BCBS 239 and EBA guidelines. These regulations emphasize the importance of being able to track data lineage, which indirectly supports the long-term retention of relevant metadata [51, 134, 81].
A2-2	Metadata	Metadata is stored in a durable and searchable system designed for long-term access, independent of the related data's lifecycle.	1	Enterprise data catalog tools (e.g., Collibra, Informatica, Alation) store metadata in centralized, searchable repositories. These systems support long-term access to metadata regardless of the availability or lifecycle status of the associated datasets [5, 30, 39].
A2-3	Metadata	Metadata records indicate when the associated data is no longer available and include a reason or status (e.g., archived, deleted).	0	While some metadata systems support status indicators, there is no clear evidence in public documentation or regulatory guidance that banks consistently use such indicators to reflect data availability. No explicit requirements or industry-wide practices were found that mandate metadata records to include such statuses.
Average Score:			**0.66**	

Interoperability

I1-1	Data & Metadata	Meta(data) is stored and exchanged in structured, machine-readable formats (e.g., JSON, XML, CSV) that support human and system interpretation.	1	In the banking sector, data is typically stored and exchanged using structured formats. Regulatory frameworks like AnaCredit require data submissions in SDMX-ML, an XML-based format that is both human and machine readable [82]. Additionally, many banks expose data via APIs using XML and JSON, which supports both automated systems and human interpretation [2, 7, 8].
I1-2	Metadata	Metadata follows standardized and widely adopted schemas (e.g., DDI) where applicable.	1	Banks often follow standardized metadata schemas depending on the context. In regulatory reporting, the SDMX standard defines metadata structures that are adopted across central banks and financial institutions [28]. Additionally, INEXDA promotes the use of a shared metadata schema for granular financial datasets, based on the DDI standard [56].
I1-3	Infrastructure	The data and metadata formats used are open, publicly documented and commonly adopted across the financial domain.	1	Formats such as SDMX-ML, JSON and XML are open and publicly documented. They are also widely adopted across the financial domain. SDMX is mandated for ECB reporting and JSON is the preferred format for APIs [62]. Metadata schemas used in SDMX and INEXDA are also open, with publicly available documentation [28, 56].
Average Score:			**1**	

I2-1	Data & Metadata	Variable names, classifications and terminology used in (meta)data conform to recognized domain vocabularies.	1	By checking vocabulary documents of common banks, such as ABN Amro, it can be concluded that banks use common financial and risk-related terms, many of which are drawn from regulatory frameworks like Basel [48]. This shows use of standardized, domain-specific terminology.
I2-2	Metadata	Vocabularies are documented, versioned and accessible to data users for interpretation and integration.	1	The vocabulary published by ABN Amro provides definitions for each term, includes a publication date and is publicly accessible [48]. This meets the criterion for documentation and accessibility.
I2-3	Metadata	Vocabularies are themselves FAIR: they use a unique and persistent identifier, are resolvable via a standardized protocol and use a formal knowledge representation language.	0	The vocabulary does document the terms, but does not show signs of doing it in a FAIR-compliant way. There is no evidence of the use of a unique identifier [48].
I2-4	Metadata	Vocabularies are available in a machine-readable format.	0	The vocabulary is in human-readable PDF format and does not follow a machine-readable structure, such as XML or JSON [48].
Average Score:			**0.5**	

I3-1	Data	Data includes explicit references to other datasets or tables, using unique keys or identifiers that support linkage.	1	AnaCredit defines explicit links between credit instruments (e.g., loans), counterparties (e.g., borrowers) and protection items (e.g., collateral). This is done through identifiers ("keys") defined in the datasets, ensuring that data on for example loans can be linked to the corresponding borrower [16].

I3-2	Metadata	Metadata defines and documents relationships between datasets, including the nature and direction of the linkage.	1	In the AnaCredit framework, metadata explicitly defines the relationships between entities. These relationships include descriptive metadata that clarify the role of each linked entity [6].
I3-3	Infrastructure	References are stored in structured, machine-readable formats.	1	In regulatory contexts such as AnaCredit, institutions are required to submit credit data using the SDMX standard, which supports structured, machine-readable data formats like SDMX-ML and JSON [28].
Average Score:			**1**	

Reusability				
R1-1	Metadata	Metadata includes essential technical attributes such as definitions, data types, units of measurement and timestamps.	1	The INEXDA metadata schema contains detailed descriptive fields such as variable name, label, unit of measure, descriptions and keywords [56]. This ensures that the metadata contains important information needed to comprehensively describe the structure and meaning of the dataset.
R1-2	Metadata	Metadata captures contextual information such as the purpose of data collection, applicability and known limitations.	1	The INEXDA schema includes fields for describing the purpose of the dataset, methodological notes and intended usage. It also allows for descriptions about the data and how it should be interpreted [56]. This supports a clear understanding of the context in which the data was produced and how it should be used.

R1-3	Metadata	Metadata includes version information and a documented change history to track modifications over time.	0	The INEXDA metadata schema does not include explicit fields for tracking metadata version numbers or maintaining a change history [56]. Although datasets may be updated, the schema does not provide a structured mechanism for documenting when changes were made, what was changed and why.
R1-4	Metadata	Metadata is sufficiently detailed to support the correct interpretation and reuse of data in different operational contexts.	1	Following the INEXDA metadata schema, the metadata provides sufficiently detailed descriptions to support correct interpretation and reuse of the data [56].
Average Score:			**0.75**	

R1.1-1	Data & Metadata	Clear usage rights and restrictions are stated for both data and metadata.	1	Banks clearly define data usage rights through privacy statements, API terms and legal agreements. For example, the processing of personal data is clearly documented, defining rights and restrictions to the use of personal data [32].
R1.1-2	Data & Metadata	Stated usage rights comply with applicable legal regulations (e.g., GDPR) and institutional data policies.	1	Data gets processed in compliance with the provisions of the GDPR and the applicable local data protection law [32]
R1.1-3	Data & Metadata	Usage terms are easy to locate, accessible and written in understandable language.	1	Data protection policies are made easily accessible on public websites, so that users and clients understand the scope and limitations of data usage [4, 23, 32]. Based on these publications, it can be reasonably assumed that similar usage rights and conditions are also established internally for users accessing data and metadata within institutional systems.
Average Score:			**1**	

R1.2-1	Metadata	Metadata describes the data's origin and the conditions under which it was collected.	1	Regulatory reporting frameworks like AnaCredit require institutions to submit information that includes identification of the source [6]. Moreover, the INEXDA's metadata schema includes fields for "Creator" and "Publication Date", which capture the origin of the data [15, 56].
R1.2-2	Metadata	Metadata documents changes to the data, including transformations, cleaning or aggregation steps.	0	There is no public or external evidence that this information is systematically captured in metadata files. The INEXDA metadata schema does not specify dedicated fields for recording transformations applied to the data [15, 56]. Without structured support for documenting data processing steps, users cannot trace how the raw data has changed over time.
R1.2-3	Metadata	Metadata includes timestamps, version information and identification of responsible individuals or systems for updates.	0	There is no systematic version control infrastructure (e.g. timestamps, version history or update contributor metadata) visible in publicly available documentation. Metadate schemas such as INEXDA's lack such fields. Therefore, this criterion is considered not fulfilled [56].
R1.2-4	Metadata	Provenance information is systematically maintained as part of the data management process.	0	No publicly available evidence was found that suggests financial institutions systematically maintain provenance records.
Average Score:			**0.25**	

R1.3-1	Data & Metadata	(Meta)data use terminology that aligns with established community conventions or domain-specific language.	1	The AnaCredit reporting guideline requires the use of standardized code lists and definitions mandated by the ECB [16]. These controlled vocabularies ensure semantic consistency and align with domain standards for financial reporting.
R1.3-2	Metadata	Metadata follows domain-recognized structures or standards.	1	As mentioned before, guideline (EU) 2017/2335 requires AnaCredit data to follow the SDMX standards [82]. This provides a structured and widely accepted format for documenting metadata.
Average Score:			**1**	

Appendix B

FAIR Assessment P2P Lending Platforms Using Open Documentation and Literature

This appendix presents the FAIRness assessment of P2P lending platforms, based on a combination of acadamic literature, regulatory frameworks, institutional guidelines and available documentation on P2P lending platforms data practices. The assessment follows he original version of the FAIR evaluation framework, as described in Chapter 5. Each FAIR sub-principle is evaluated across specific criteria. Each criterion is scored either non-compliant (0) or compliant (1). The average per sub-principle represents the degree of alignment with the FAIR principles.

The assessment focuses mainly on the platform Bondora as a representative case study of European P2P lending platforms. This choice is based on Bondora's scale, regulatory status and public accessibility of its documentation and API. These sources are used to assess the platform's data governance practices. Where Bondora's documentation is limited, the assessment is supplemented with insights from other P2P lending platforms.

Regarding metadata practices no publicly available information was found for Bondora or comparable platforms. However, based on industry best practices and regulatory expectations, it is highly likely that structured internal metadata exists to support data management practices. As public validation is not possible, this analysis uses the European Single Access Point (ESAP) Regulation (EU 2023/2859) as a reference for expected metadata practices. ESAP is a quite new regulatory initiative aimed at centralising access to publicly disclosed financial information across the EU [26]. While P2P lending platforms are not currently within ESAP's scope, the regulation is explicitly forward-looking and allows for future inclusion of additional financial actors. Furthermore, P2P platforms are already regulated under the Crowdfunding Regulation (EU 2020/1503) and are going to be included in the upcoming Financial Data Access (FIDA) Regulation [34, 25], reinforcing their growing role in the EU financial data infrastructure. Taken together, these developments support the view that P2P lending platforms are likely candidates for future inclusion in ESAP. Therefore, The ESAP framework is used in this assessment to approximate internal metadata practices.

TABLE B.1: FAIR Assessment P2P Lending Platforms Using Open Documentation and Literature

ID	**Target**	**Criterion**	**Score**	**Justification**
Findability				
F1-1	Data	Each data object is assigned an identifier that is unique at the institutional level. Where possible, globally unique identifiers, like the Legal Entity Identifier (LEI), are used.	1	P2P lending platforms typically assign unique identifiers to each data object, such as loans and users, within their systems [118, 1]. The Bondora public API also illustrates this approach, providing distinct identifiers for each entity [11]. These identifiers are unique within the platform's ecosystem, enabling reliable internal referencing even if they are not globally unique.
F1-2	Data	Identifiers remain stable and unchanged throughout the data lifecycle, this includes among others system migrations, updates and archiving.	1	P2P lending platforms assign stable, persistent identifiers to track data entities across their lifecycle. For example Bondora uses such identifiers in their API responses [11].
F1-3	Metadata	Each metadata record is assigned an identifier that is unique within the institution.	1	Although public documentation is limited for many P2P lending platforms, this does not preclude the existence of internal mechanisms for uniquely identifying metadata. Evidence from P2P lending platforms such as Funding Societies, which use a structured data management system such as Atlan that assigns unique identifiers to metadata, supports the conclusion that such practices are present in this sector [9, 113]. Therefore, it is concluded that this criterion is met.

F1-4	Metadata	Identifiers remain stable and unchanged throughout the data lifecycle, this includes among others system migrations, updates and archiving.	1	Some P2P lending platforms, like Funding Societies, use structured data governance tools, such as Atlan, which support the stability of metadata identifiers across the data lifecycle [9, 113]. Based on the presence of such practices in this sector, it is concluded that this criterion is met.
Average Score:			**1**	

F2-1	Metadata	Metadata includes core descriptive elements hat support effective search and discovery, such as title, description, keywords, format and classification terms.	1	P2P lending platforms such as Bondora provide access to their data through APIs, which are described using Swagger [11]. This documentation provides descriptive elements such as naming, format, data types and descriptions [11]. While not equivalent to rich metadata, the schema fulfills the requirement for basic descriptive metadata that enhances findability.
F2-2	Metadata	Metadata includes basic provenance elements (e.g., origin, creator, creation date) that support findability.	1	While limited public information about metadata practices could be found, the ESAP regulation can be used here as representative for likely internal metadata practices. The ESAP Regulation prescribes several provenance related fields for metadata, particularly for the purpose of searchability [26]. Therefore it is concluded that this criterion is met.
Average Score:			**1**	

F3-1	Metadata	Metadata includes an explicit reference to the corresponding dataset using a unique and stable identifier.	1	ESAP Regulation requires that the metadata should provide a stable link to the data it references to [35]. Assuming the ESAP Regulation aligns with internal practices, it is reasonable to conclude that datasets would be linked through stable references.
F3-2	Metadata	The linkage between metadata and data is represented in a structured, machine-readable format.	1	Although Bondora's public API does not expose metadata files in a formal schema, the API responses are delivered in machine-readable formats such as JSON and XML [11]. Furthermore, under ESAP Regulation, metadata must be machine-readable and linked to the data content [26]. Therefore, it is highly likely that the linkage between metadata and data is represented in a structured and machine-readable format.
Average Score:			**1**	

F4-1	Data	Data is registered/indexed in a searchable resource.	1	P2P lending platforms that use structured data governance tools, such as Atlan, can index datasets in searchable catalogs [9]. This allows users to discover and access data relevant to credit risk assessment.
F4-2	Metadata	Metadata is registered/indexed in a searchable resource.	1	Platforms using tools like Atlan index metadata alongside datasets, making the metadata searchable [9]. This allows users to find metadata independently or in combination with data.

F4-3	Infrastructure	The searchable resource supports keyword or attribute-based search across datasets and metadata records.	1	Data catalog systems, such as Atlan, adopted by P2P lending platforms support specific filtering options to search based on specific keywords or attributes, enabling users to locate datasets and metadata through filters [9].
F4-4	Infrastructure	The searchable resource provides comprehensive coverage of all relevant datasets and metadata.	0	While data catalogs like Atlan can index all relevant content, the extent to which this is done is dependent on the P2P lending platform [9]. Given the lack of public documentation from P2P lending platforms and the sector's variability in data governance maturity, it is concluded that it is unlikely that all relevant datasets and metadata are consistently indexed.
F4-5	Infrastructure	The searchable resource provides a machine-readable interface for automated search and discovery.	1	Catalog tools like Atlan provide machine-readable interfaces via APIs, allowing automated querying of indexed metadata and datasets [9].
Average Score:			**0.8**	

Accessibility				
A1-1	Data	Data can be retrieved using their identifier via a communication protocol (e.g., HTTPS or FTP).	1	Bondora and similar P2P platforms expose their data through standardized RESTful APIs over HTTPS [11, 131]. These APIs allow retrieval of data objects (e.g., loans) using unique identifiers like loanId, fulfilling the FAIR requirement that data be retrievable via a standard, open communication protocol.

A1-2	Metadata	Metadata is retrievable using its identifier via the same communication protocol as the associated data.	1	In platforms like Bondora, metadata is retrievable via the same RESTful API used for data access [11]. While metadata is embedded rather than provided as a separate schema file, it is still accessible via the same standardized protocol. Generally, if a RESTful API allows for data retrieval via identifiers, it also provides access to the associated metadata [159].
A1-3	Infrastructure	The communication protocol used to access (meta)data is well-documented and based on established standards.	1	Data and metadata retrieval occur over HTTPS, which is a globally standardized, documented and widely implemented communication protocol [68].
Average Score:			**1**	

A1.1-1	Infrastructure	Data retrieval relies on open, non-proprietary communication protocols that are freely and universally implementable.	1	Bondora's RESTful API is built on HTTPS, an open and non-proprietary protocol that is freely implementable by any system [128]. This ensures universal accessibility without the need for vendor-specific software or proprietary tools.
A1.1-2	Infrastructure	The communication protocols used are publicly documented and accessible.	1	HTTPS is a globally standardized communication protocol and publicly documented [85]. Bondora's use of HTTPS ensures that the communication layer is transparent and fully accessible to any developer [11].
A1.1-3	Infrastructure	The same protocol is consistently applied across systems and departments within the institution.	1	Bondora consistently uses HTTPS as the only protocol across its web interface and API infrastructure [11].

A1.1-4	Infrastructure	The protocol allows for automated data retrieval and does not rely on manual human intervention.	1	HTTPS supports programmatic data access without human involvement, provided the system is authenticated. Bondora's use of HTTPS in combination with OAuth2 authentication and JSON-formatted responses enables automated retrieval of (meta)data [11].
Average Score:			**1**	

A1.2-1	Infrastructure	The communication protocol supports authentication to verify the identity of users or systems accessing (meta)data.	1	Bondora uses OAuth 2.0 to authenticate API clients, ensuring that only authorized users can access protected data endpoints [11]. OAuth 2.0 is a widely adopted, secure and standardized authentication framework [87].
A1.2-2	Infrastructure	The protocol supports authorization mechanisms to control access to (meta)data.	1	The platform has access control systems implemented by assigning scopes, which are specific permissions that you can request from users. This allows the platform to restrict access to specific data operations [11].
A1.2-3	Infrastructure	Access control mechanisms comply with relevant legal, regulatory and institutional requirements	1	While P2P lending platforms typically face fewer sector-specific regulations than traditional banks, they are nonetheless subject to overarching data protection and privacy laws, such as the GDPR [73]. P2P lending platforms, like Bondora, explicitly state their compliance with these frameworks [29].
Average Score:			**1**	

A2-1	Metadata	Metadata remains accessible through the same infrastructure, even after the associated dataset is archived or deleted.	1	P2P lending platforms that adopt data governance tools like Atlan can preserve metadata independently of the associated data. These systems are designed to retain metadata even when datasets are deleted or archived [9]. Although there is no public confirmation that this capability is actively use, the presence of this functionality suggests its application in environments where such tools are implemented.
A2-2	Metadata	Metadata is stored in a durable and searchable system designed for long-term access, independent of the related data's lifecycle.	1	Atlan is a durable and searchable platform to store metadata. Systems like this are designed for long-term access and ensure that metadata remains available regardless of dataset availability [9].
A2-3	Metadata	Metadata records indicate when the associated data is no longer available and include a reason or status (e.g., archived, deleted).	0	While data catalog platforms like Atlan can support statuses for when data is no longer available, it is unclear whether P2P lending platforms consistently implement such detailed metadata practices.
Average Score:			**0.66**	

Interoperability				
I1-1	Data & Metadata	Meta(data) is stored and exchanged in structured, machine-readable formats (e.g., JSON, XML, CSV) that support human and system interpretation.	1	P2P lending platforms commonly use structured, machine-readable formats such as JSON and XML. For example, the Bondora API supports both JSON and XML for data exchange [11].

I1-2	Metadata	Metadata follows standardized and widely adopted schemas (e.g., DDI) where applicable.	0	The ESAP Regulation requires metadata to be structured and machine-readable, often through formats like XML [35]. However, it does not mandate the use of domain-recognized metadata schemas such as DDI. There is no public evidence that Bondora or similar P2P lending platforms adopt such standardized schemas. Therefore, this criterion is not met.
I1-3	Infrastructure	The data and metadata formats used are open, publicly documented and commonly adopted across the financial domain.	1	ESAP compliant formats such as JSON and XBRL are open, publicly documented and widely adopted across the financial industry. Given Bondora's integration with EU regulatory ecosystems and its use of JSON and XML in its API, it is reasonable to assume that both data and metadata are maintained in similarly formats [11, 35].
Average Score:			**0.66**	

I2-1	Data & Metadata	Variable names, classifications and terminology used in (meta)data conform to recognized domain vocabularies.	1	Bondora's API uses common financial terminology and abbreviations and it also shows alignment with international code lists such as ISO 3166-1 A-2 country codes and ISO 8601 data and time format [11]. This shows the alignment with domain accepted vocabularies.
I2-2	Metadata	Vocabularies are documented, versioned and accessible to data users for interpretation and integration.	0	Although Bondora's API includes field-level descriptions and enumerated values, these are not maintained as formal vocabularies with independent documentation [11]. No Evidence could be found of a formally documented vocabulary. As a result this criterion is not met.

I2-3	Metadata	Vocabularies are themselves FAIR: they use a unique and persistent identifier, are resolvable via a standardized protocol and use a formal knowledge representation language.	0	As far as can be determined from publicly available information, P2P lending platforms such as Bondora do not use formal vocabularies. Since no evidence of defined vocabularies was found, it follows that this criterion is also not met.
I2-4	Metadata	Vocabularies are available in a machine-readable format.	0	While Bondora's API uses structured formats such as JSON and XML to represent data [11], the lack of formal terminology also makes it not machine-readable.
Average Score:			**0.25**	

I3-1	Data	Data includes explicit references to other datasets or tables, using unique keys or identifiers that support linkage.	1	P2P lending platforms commonly structure their data to include internal references between related entities. These references are typically implemented through identifiers or foreign keys, enabling the connection of related datasets [1].
I3-2	Metadata	Metadata defines and documents relationships between datasets, including the nature and direction of the linkage.	0	No public metadata describing the nature of links between datasets could be found. Furthermore, ESAP does not explicitly require qualified descriptions of dataset relationships [26]. As a result, this criterion is not fulfilled.
I3-3	Infrastructure	References are stored in structured, machine-readable formats.	1	Although Bondora's public API does not expose metadata files in a formal schema, the API responses are delivered in machine-readable formats such as JSON and XML [11]. Furthermore, under ESAP Regulation, metadata must be machine-readable [26]. Therefore, it is highly likely that the references are stored in structured and machine-readable formats.
Average Score:			**0.66**	

Reusability				
R1-1	Metadata	Metadata includes essential technical attributes such as definitions, data types, units of measurement and timestamps.	1	ESAP requires structured, machine-readable metadata including among others name, data type, data reference, legal framework and versioning [26, 35]. Although not a complete technical specification, these core elements are sufficient to support basic technical interpretation. Therefore, this criterion is met.
R1-2	Metadata	Metadata captures contextual information such as the purpose of data collection, applicability and known limitations.	0	ESAP focuses on identification and discoverability, but does not require metadata to describe the purpose, applicability or limitations of the data [26, 35]. The focus of ESAP is not reuse of data. Since, no other public documentation about metadata for P2P lending platforms could be found this criterion is not met.
R1-3	Metadata	Metadata includes version information and a documented change history to track modifications over time.	0	The ESAP Regulation requires metadata to include version information, however it does not mandate a documented change history [26, 35]. Further, there is no indication that Bondora or other P2P lending platforms implement such detailed historical tracking in their metadata. Thus, this criterion is not met.
R1-4	Metadata	Metadata is sufficiently detailed to support the correct interpretation and reuse of data in different operational contexts.	0	ESAP supports discoverability and identification but does not require metadata that fully supports reuse [26, 35]. No supplementary metadata from P2P lending platforms like Bondora was found to fill these gaps. As a result, this criterion is not met.
Average Score:			**0.25**	

R1.1-1	Data & Metadata	Clear usage rights and restrictions are stated for both data and metadata.	1	Although internal usage rights are not publicly avialable, Bondora's Terms of Use and Privacy Policy explicitly describes the usage rights and restrictions for users, including data handling practices and user obligations. This documentation provides clear information on how data can be used and any associated limitations [29]. Therefore, it can be assumed that internal usage rights are also clearly stated.
R1.1-2	Data & Metadata	Stated usage rights comply with applicable legal regulations (e.g., GDPR) and institutional data policies.	1	Bondora's Terms of Use and Privacy Policy clearly state compliance with legal requirements, such as the GDPR [29]. This confirms that its data usage policies adhere to legally binding standards
R1.1-3	Data & Metadata	Usage terms are easy to locate, accessible and written in understandable language.	1	The Terms of Use and Privacy Policy of Bondora are easily accessible on Bondora's website [29]. They are presented in a clear and understandable manner so that users can easily understand them.
Average Score:			**1**	

R1.2-1	Metadata	Metadata describes the data's origin and the conditions under which it was collected.	0	The ESAP Regulation does not require metadata to include information about data origin or collection conditions [26, 35]. No additional documentation from Bondora or similar platforms provides such provenance information. As a result, this criterion is not fulfilled.

R1.2-2	Metadata	Metadata documents changes to the data, including transformations, cleaning or aggregation steps.	0	The ESAP framework does not mandate that metadata include details on data processing activities such as transformations, cleaning, or aggregations [26, 35]. Moreover, Bondora provides no public information on change tracking. Although version control can be done internally, this is not specified in regulations or publicly documented by a P2P lending platform.
R1.2-3	Metadata	Metadata includes timestamps, version information and identification of responsible individuals or systems for updates.	0	While submission timestamps and versioning are described by ESAP, it does not require identification of the individuals responsible for those updates [26, 35]. Since only part of the criterion is met, it is scored as non-compliant.
R1.2-4	Metadata	Provenance information is systematically maintained as part of the data management process.	0	The ESAP Regulation does not define requirements for systematic provenance data management [26, 35]. There is no indication that Bondora or similar platforms provide this metadata publicly. Therefore, this criterion is not fulfilled.
Average Score:			**0**	

R1.3-1	Data & Metadata	(Meta)data use terminology that aligns with established community conventions or domain-specific language.	1	Bondora's API makes use of certain standardized code lists, such as ISO 3166-1 A-2 country codes and ISO 8601 data and time format [11]. This demonstrates that domain conventions are being met.
R1.3-2	Metadata	Metadata follows domain-recognized structures or standards.	1	By aligning with the ESAP Regulation, metadata is expected to follow structured and domain-recognized formats suitable for financial and regulatory contexts [26].
Average Score:			**1**	

Appendix C

Survey Response Traditional Financial Institution

This appendix contains the response provided by a representative of a traditional financial institution to the FAIR alignment assessment survey. The survey was designed to evaluate the institution's data management practices in relation to the FAIR principles. The response serves to complement and validate the literature based analysis conducted in the main body of this thesis. For consistency and confidentiality, identifying details have been anonymized.

Table C.1: FAIR Assessment Survey Response Traditional Financial Institution

ID	Target	Criterion	Score
Findability			
F1-1	Data	Each data object is assigned an identifier that is unique at the institutional level. Where possible, globally unique identifiers, like the Legal Entity Identifier (LEI), are used.	0
F1-2	Data	Identifiers remain stable and unchanged throughout the data lifecycle, this includes among others system migrations, updates and archiving.	0
F1-3	Metadata	Each metadata record is assigned an identifier that is unique within the institution.	1
F1-4	Metadata	Identifiers remain stable and unchanged throughout the data lifecycle, this includes among others system migrations, updates and archiving.	1
Average Score:			**0.5**

F2-1	Metadata	Metadata includes core descriptive elements hat support effective search and discovery, such as title, description, keywords, format and classification terms.	1
F2-2	Metadata	Metadata includes basic provenance elements (e.g., origin, creator, creation date) that support findability.	1
Average Score:			**1**

F3-1	Metadata	Metadata includes an explicit reference to the corresponding dataset using a unique and stable identifier.	1
F3-2	Metadata	The linkage between metadata and data is represented in a structured, machine-readable format.	1
Average Score:			**1**

F4-1	Data	Data is registered/indexed in a searchable resource.	1
F4-2	Metadata	Metadata is registered/indexed in a searchable resource.	1
F4-3	Infrastructure	The searchable resource supports keyword or attribute-based search across datasets and metadata records.	1

F4-4	Infrastructure	The searchable resource provides comprehensive coverage of all relevant datasets and metadata.	0
F4-5	Infrastructure	The searchable resource provides a machine-readable interface for automated search and discovery.	0
Average Score:			**0.6**

Accessibility			
A1-1	Data	Data can be retrieved using their identifier via a communication protocol (e.g., HTTPS or FTP).	0
A1-2	Metadata	Metadata is retrievable using its identifier via the same communication protocol as the associated data.	0
A1-3	Infrastructure	The communication protocol used to access (meta)data is well-documented and based on established standards.	1
Average Score:			**0.33**

A1.1-1	Infrastructure	Data retrieval relies on open, non-proprietary communication protocols that are freely and universally implementable.	0
A1.1-2	Infrastructure	The communication protocols used are publicly documented and accessible.	0
A1.1-3	Infrastructure	The same protocol is consistently applied across systems and departments within the institution.	0
A1.1-4	Infrastructure	The protocol allows for automated data retrieval and does not rely on manual human intervention.	1
Average Score:			**0.25**

A1.2-1	Infrastructure	The communication protocol supports authentication to verify the identity of users or systems accessing (meta)data.	0
A1.2-2	Infrastructure	The protocol supports authorization mechanisms to control access to (meta)data.	0

A1.2-3	Infrastructure	Access control mechanisms comply with relevant legal, regulatory and institutional requirements	1
Average Score:			**0.33**

A2-1	Metadata	Metadata remains accessible through the same infrastructure, even after the associated dataset is archived or deleted.	0
A2-2	Metadata	Metadata is stored in a durable and searchable system designed for long-term access, independent of the related data's lifecycle.	0
A2-3	Metadata	Metadata records indicate when the associated data is no longer available and include a reason or status (e.g., archived, deleted).	1
Average Score:			**0.33**

Interoperability			
I1-1	Data & Metadata	Meta(data) is stored and exchanged in structured, machine-readable formats (e.g., JSON, XML, CSV) that support human and system interpretation.	1
I1-2	Metadata	Metadata follows standardized and widely adopted schemas (e.g., DDI) where applicable.	1
I1-3	Infrastructure	The data and metadata formats used are open, publicly documented and commonly adopted across the financial domain.	0
Average Score:			**0.66**

I2-1	Data & Metadata	Variable names, classifications and terminology used in (meta)data conform to recognized domain vocabularies.	1
I2-2	Metadata	Vocabularies are documented, versioned and accessible to data users for interpretation and integration.	1
I2-3	Metadata	Vocabularies are themselves FAIR: they use a unique and persistent identifier, are resolvable via a standardized protocol and use a formal knowledge representation language.	0
I2-4	Metadata	Vocabularies are available in a machine-readable format.	1
Average Score:			**0.75**

I3-1	Data	Data includes explicit references to other datasets or tables, using unique keys or identifiers that support linkage.	0
I3-2	Metadata	Metadata defines and documents relationships between datasets, including the nature and direction of the linkage.	1
I3-3	Infrastructure	References are stored in structured, machine-readable formats.	0
Average Score:			**0.33**
Reusability			
R1-1	Metadata	Metadata includes essential technical attributes such as definitions, data types, units of measurement and timestamps.	1
R1-2	Metadata	Metadata captures contextual information such as the purpose of data collection, applicability and known limitations.	0
R1-3	Metadata	Metadata includes version information and a documented change history to track modifications over time.	1
R1-4	Metadata	Metadata is sufficiently detailed to support the correct interpretation and reuse of data in different operational contexts.	1
Average Score:			**0.75**
R1.1-1	Data & Metadata	Clear usage rights and restrictions are stated for both data and metadata.	1
R1.1-2	Data & Metadata	Stated usage rights comply with applicable legal regulations (e.g., GDPR) and institutional data policies.	1
R1.1-3	Data & Metadata	Usage terms are easy to locate, accessible and written in understandable language.	0
Average Score:			**0.66**
R1.2-1	Metadata	Metadata describes the data's origin and the conditions under which it was collected.	0

R1.2-2	Metadata	Metadata documents changes to the data, including transformations, cleaning or aggregation steps.	1
R1.2-3	Metadata	Metadata includes timestamps, version information and identification of responsible individuals or systems for updates.	1
R1.2-4	Metadata	Provenance information is systematically maintained as part of the data management process.	0
Average Score:			**0.5**

R1.3-1	Data & Metadata	(Meta)data use terminology that aligns with established community conventions or domain-specific language.	1
R1.3-2	Metadata	Metadata follows domain-recognized structures or standards.	0
Average Score:			**0.5**

Appendix D

Survey Response Alternative Finance Institution

This appendix contains the response provided by a representative of a alternative financial institution to the FAIR alignment assessment survey. The survey was designed to evaluate the institution's data management practices in relation to the FAIR principles. The response serves to complement and validate the literature based analysis conducted in the main body of this thesis. For consistency and confidentiality, identifying details have been anonymized.

TABLE D.1: FAIR Assessment Survey Response Alternative Finance Institution

ID	Target	Criterion	Score
Findability			
F1-1	Data	Each data object is assigned an identifier that is unique at the institutional level. Where possible, globally unique identifiers, like the Legal Entity Identifier (LEI), are used.	1
F1-2	Data	Identifiers remain stable and unchanged throughout the data lifecycle, this includes among others system migrations, updates and archiving.	1
F1-3	Metadata	Each metadata record is assigned an identifier that is unique within the institution.	1
F1-4	Metadata	Identifiers remain stable and unchanged throughout the data lifecycle, this includes among others system migrations, updates and archiving.	0
Average Score:			**0.75**
F2-1	Metadata	Metadata includes core descriptive elements hat support effective search and discovery, such as title, description, keywords, format and classification terms.	1
F2-2	Metadata	Metadata includes basic provenance elements (e.g., origin, creator, creation date) that support findability.	1
Average Score:			**1**
F3-1	Metadata	Metadata includes an explicit reference to the corresponding dataset using a unique and stable identifier.	1
F3-2	Metadata	The linkage between metadata and data is represented in a structured, machine-readable format.	1
Average Score:			**1**
F4-1	Data	Data is registered/indexed in a searchable resource.	1
F4-2	Metadata	Metadata is registered/indexed in a searchable resource.	1
F4-3	Infrastructure	The searchable resource supports keyword or attribute-based search across datasets and metadata records.	1

F4-4	Infrastructure	The searchable resource provides comprehensive coverage of all relevant datasets and metadata.	1
F4-5	Infrastructure	The searchable resource provides a machine-readable interface for automated search and discovery.	0
Average Score:			**0.8**

Accessibility			
A1-1	Data	Data can be retrieved using their identifier via a communication protocol (e.g., HTTPS or FTP).	1
A1-2	Metadata	Metadata is retrievable using its identifier via the same communication protocol as the associated data.	1
A1-3	Infrastructure	The communication protocol used to access (meta)data is well-documented and based on established standards.	0
Average Score:			**0.66**

A1.1-1	Infrastructure	Data retrieval relies on open, non-proprietary communication protocols that are freely and universally implementable.	1
A1.1-2	Infrastructure	The communication protocols used are publicly documented and accessible.	1
A1.1-3	Infrastructure	The same protocol is consistently applied across systems and departments within the institution.	0
A1.1-4	Infrastructure	The protocol allows for automated data retrieval and does not rely on manual human intervention.	0
Average Score:			**0.5**

A1.2-1	Infrastructure	The communication protocol supports authentication to verify the identity of users or systems accessing (meta)data.	1
A1.2-2	Infrastructure	The protocol supports authorization mechanisms to control access to (meta)data.	1

A1.2-3	Infrastructure	Access control mechanisms comply with relevant legal, regulatory and institutional requirements	1
Average Score:			**1**

A2-1	Metadata	Metadata remains accessible through the same infrastructure, even after the associated dataset is archived or deleted.	1
A2-2	Metadata	Metadata is stored in a durable and searchable system designed for long-term access, independent of the related data's lifecycle.	1
A2-3	Metadata	Metadata records indicate when the associated data is no longer available and include a reason or status (e.g., archived, deleted).	0
Average Score:			**0.66**

Interoperability			
I1-1	Data & Metadata	Meta(data) is stored and exchanged in structured, machine-readable formats (e.g., JSON, XML, CSV) that support human and system interpretation.	1
I1-2	Metadata	Metadata follows standardized and widely adopted schemas (e.g., DDI) where applicable.	1
I1-3	Infrastructure	The data and metadata formats used are open, publicly documented and commonly adopted across the financial domain.	1
Average Score:			**1**

I2-1	Data & Metadata	Variable names, classifications and terminology used in (meta)data conform to recognized domain vocabularies.	1
I2-2	Metadata	Vocabularies are documented, versioned and accessible to data users for interpretation and integration.	1
I2-3	Metadata	Vocabularies are themselves FAIR: they use a unique and persistent identifier, are resolvable via a standardized protocol and use a formal knowledge representation language.	1
I2-4	Metadata	Vocabularies are available in a machine-readable format.	0
Average Score:			**0.75**

I3-1	Data	Data includes explicit references to other datasets or tables, using unique keys or identifiers that support linkage.	1
I3-2	Metadata	Metadata defines and documents relationships between datasets, including the nature and direction of the linkage.	1
I3-3	Infrastructure	References are stored in structured, machine-readable formats.	1
Average Score:			**1**

Reusability			
R1-1	Metadata	Metadata includes essential technical attributes such as definitions, data types, units of measurement and timestamps.	1
R1-2	Metadata	Metadata captures contextual information such as the purpose of data collection, applicability and known limitations.	1
R1-3	Metadata	Metadata includes version information and a documented change history to track modifications over time.	0
R1-4	Metadata	Metadata is sufficiently detailed to support the correct interpretation and reuse of data in different operational contexts.	1
Average Score:			**0.75**

R1.1-1	Data & Metadata	Clear usage rights and restrictions are stated for both data and metadata.	1
R1.1-2	Data & Metadata	Stated usage rights comply with applicable legal regulations (e.g., GDPR) and institutional data policies.	1
R1.1-3	Data & Metadata	Usage terms are easy to locate, accessible and written in understandable language.	0
Average Score:			**0.66**

R1.2-1	Metadata	Metadata describes the data's origin and the conditions under which it was collected.	1

R1.2-2	Metadata	Metadata documents changes to the data, including transformations, cleaning or aggregation steps.	1
R1.2-3	Metadata	Metadata includes timestamps, version information and identification of responsible individuals or systems for updates.	0
R1.2-4	Metadata	Provenance information is systematically maintained as part of the data management process.	0
Average Score:			**0.5**

R1.3-1	Data & Metadata	(Meta)data use terminology that aligns with established community conventions or domain-specific language.	1
R1.3-2	Metadata	Metadata follows domain-recognized structures or standards.	1
Average Score:			**1**