University of Zurich[UZH]

MASTER THESIS – Communication Systems Group, Prof. Dr. Burkhard Stiller

Prioritization of Quality Requirements in Decentralized Identity Applications

Venusan Velrajah
Zurich, Switzerland
Student ID: 17-706-706

Supervisor: Daria Schumm, Thomas Grübl, Prof. Dr. Burkhard Stiller
Date of Submission: November 14, 2025

University of Zurich
Department of Informatics (IFI)
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland

ifi

Master Thesis
Communication Systems Group (CSG)
Department of Informatics (IFI)
University of Zurich
Binzmuehlestrasse 14, CH-8050 Zurich, Switzerland
URL: http://www.csg.uzh.ch/

Declaration of Independence

I hereby declare that I have composed this work independently and without the use of any aids other than those declared (including generative AI such as ChatGPT). I am aware that I take full responsibility for the scientific character of the submitted text myself, even if AI aids were used and declared (after written confirmation by the supervising professor). All passages taken verbatim or in sense from published or unpublished writings are identified as such. The work has not yet been submitted in the same or similar form or in excerpts as part of another examination.

Zürich, 14.11.2025

Signature of student

Abstract

In der sich ständig weiterentwickelnden Landschaft des digitalen Identitätsmanagements stellen dezentrale und Self-Sovereign Identitätssysteme (DI/SSI) einen Paradigmenwechsel hin zu einer benutzerzentrierten Kontrolle dar, der eine Neubewertung der Definition und Priorisierung von Qualitätsanforderungen durch alle Systemakteure erforderlich macht. Diese Arbeit treibt das Requirements Engineering (RE) für DI/SSI-Systeme voran, indem sie nicht-funktionale Anforderungen (NFRs) aus den unterschiedlichen Perspektiven von Identitätsinhabern, Ausstellern und Prüfern priorisiert, über den typischen organisationszentrierten Ansatz hinausgeht und nutzerzentrierte Ziele wie Kontrolle, Datenschutz und Vertrauen berücksichtigt. Die Studie erstellt stakeholder-spezifische NFR-Rankings und untersucht, inwiefern konkrete Systemfunktionen als unterstützend für diese Eigenschaften wahrgenommen werden. Unter Verwendung eines strukturierten Verantwortungsschemas und einer rollenbasierten Zuordnung stellt diese Arbeit die Software Quality Requirements Importance (SQRI)-Skala vor und wendet die Best-Worst Scaling (BWS) an, eine in diesem Zusammenhang neuartige Priorisierungstechnik, die in klaren, szenariobasierten Umfragen für jede Interessengruppe zum Einsatz kommt. Die Instrumente, die auf die einzelnen Interessengruppen zugeschnitten sind, in Form von Umfragen sowohl in englischer als auch in deutscher Sprache verfügbar sind, ermöglichen eine fundierte, interessengruppenspezifische Bewertung der NFRs für DI/SSI-Systeme, verbessern die aktuellen Methoden des Requirements Engineering und unterstützen eine differenzierte Analyse der Prioritätsmuster bei Identitätsinhabern, Ausstellern und Prüfern. Die Studie liefert rangierte Sets von NFRs für jede Rolle und bietet neue Einblicke in die Zuordnung von Funktionalität und Qualität anhand von Priorisierungsmatrizen, die *Schutz*, *Authentizität*, *Sicherheit* und andere NFRs mit früheren Klassifizierungen vergleichen und quadrantenbasierte Rollenunterschiede hervorheben. Die Ergebnisse bieten umsetzbare Leitlinien für die Gestaltung von DI/SSI-Systemen und tragen zur Erweiterung der Literatur zum Requirements Engineering bei, indem sie die Prioritäten der Stakeholder und die Übereinstimmung zwischen den Erwartungen der Nutzer und den Systemfunktionen verdeutlichen.

In the evolving landscape of digital identity management, Decentralized and Self-Sovereign Identity (DI/SSI) systems represent a paradigm shift toward user-centered control, prompting a re-evaluation of how quality requirements are defined and prioritized across system stakeholders. This thesis advances Requirements Engineering (RE) for DI/SSI systems by prioritizing Non-Functional Requirements (NFRs) through the distinct perspectives of identity holders, issuers, and verifiers, moving beyond the typical organization-centric approach and addressing user-centric goals such as control, privacy, and trust. The study establishes stakeholder-specific NFR rankings and investigates how concrete system functionalities are perceived as supporting these qualities. Employing a structured responsibility scheme and role-based mapping, this thesis introduces the Software Quality Requirements Importance (SQRI) scale and applies Best-Worst Scaling (BWS), a novel prioritization technique in this context, within clear, scenario-based surveys designed for each stakeholder group. The instruments, tailored to each stakeholder as surveys and available in both English and German, enable robust, stakeholder-specific assessment of NFRs for DI/SSI systems, advancing current requirements engineering methods and supporting nuanced analysis of priority patterns across identity holders, issuers, and verifiers. The research delivers ranked sets of NFRs for each role, providing new insights into functionality-quality mappings via prioritization matrices that compare *Protection*, *Authenticity*, *Security*, and other NFRs to prior classifications and highlight quadrant-based role differences. The findings offer actionable guidance for DI/SSI system design and contribute to the broader Requirements Engineering literature by clarifying stakeholder priorities and the alignment between user expectations and system features.

Acknowledgments

My sincere thanks go to my supervisor, Daria Schumm, for her consistent support and expert guidance. Her constructive comments, careful review, and regular meetings were instrumental in shaping the study design, the survey instruments, and the final report.

I also thank Prof. Dr. Stiller and the Communication Systems Group at the University of Zurich for the opportunity and support to pursue work in Decentralized and Self-Sovereign Identity and for offering an excellent environment in which to conduct it.

I gratefully acknowledge my family, my girlfriend, and my friends for their steady encouragement and patience throughout this work. Their understanding and support were vital to its completion.

Contents

Chapter 1

Introduction

In traditional Requirements Engineering (RE), projects typically begin by gathering and defining Functional Requirements (FRs) and Non-Functional Requirements (NFRs) that align with organizational goals. While this approach works well for business-driven software, it is less effective for decentralized, user-centric systems. For instance, Decentralized Identity (DI) and Self-Sovereign Identity (SSI) shift control from centralized authorities to the individuals and organizations that utilize the system. Their focus emphasizes user agency, privacy, and trust, rather than solely institutional needs [59].

These ecosystems revolve around three main roles (identity holder (user), issuers, and verifiers) who collectively establish trust without depending on a single authority [78]. Importantly, the principles defining SSI, such as control over personal data, minimal disclosure, and security, can be expressed as quality attributes or NFRs [98]. Prior research has identified a wide range of these qualities, such as *Privacy*, *Interoperability*, *Transparency*, and *Recoverability*, as essential for constructing trustworthy digital identity systems [106].

This thesis builds on that foundation by prioritizing these requirements through empirical investigation: examining how identity holders (users), issuers, and verifiers rank the importance of various NFRs and whether they recognize how specific system functionalities support these quality attributes. By anchoring the study in established classifications of SSI properties [98], this research aims to align user-centric principles with the actual perceptions and needs of the involved stakeholders.

1.1 Motivation

Digital identity management has become a foundational element of contemporary digital services. However, existing systems continue to face challenges in meeting the increasing demands for privacy, security, and user trust [93, 104]. Traditional identity infrastructures, which depend on centralized authorities and isolated databases, leave individuals vulnerable to risks such as identity theft, data breaches, and the erosion of control over personal information [18, 32]. In response to these issues, DI, and more specifically, SSI,

has emerged as a transformative approach, placing users at the forefront of identity management. SSI prioritizes autonomy, privacy, and user control, allowing individuals to manage verifiable credentials through cryptographic methods and distributed ledgers, all without depending on a single intermediary [59, 64]. Essentially, SSI aims to address the limitations of centralized identity models while enhancing trust and privacy across various sectors, including e-government, finance, and healthcare [76, 88].

The objectives of SSI systems are closely linked to NFRs, which include aspects such as *Ownership and Control*, *Privacy and Minimal Disclosure*, *Security* and *Protection*, *Interoperability*, and *Usability* [97, 98]. Unlike FRs, which define what a system does, NFRs focus on how it should operate, for instance, securely, reliably, and transparently. Recent studies have documented SSI design properties and confirmed that experts regard qualities like *Security*, *Privacy*, and *Verifiability* as essential, while also ranking *Usability* and *Interoperability* as highly desirable [10, 98]. These findings underscored the need to address a broad spectrum of qualities to ensure SSI systems are both trustworthy and adoptable.

An important question remains: whose priorities are reflected in various classifications? Much of the literature on requirements prioritization has concentrated on organizational perspectives, aiming to optimize delivery within budgetary or time constraints [3, 42]. Techniques such as AHP, MoSCoW, and cost–value frameworks frequently rely on product managers and engineers rather than considering the end-users themselves [46, 57]. This organization-centric focus risks neglecting the viewpoints of those whose trust and adoption are vital for the success of SSI. Research in software quality consistently emphasizes that if user priorities are not adequately captured, systems may fail despite being technically sound [68, 83]). This gap is particularly pressing in the context of SSI, where the system's legitimacy hinges on user acceptance, issuer compliance, and verifier trust [21, 37]).

This is where the current thesis comes into play. To design genuinely user-centric systems, it is essential to move beyond organizational interests and gain a deeper understanding of which qualities matter most to identity holders (users), issuers, and verifiers, the three key roles in SSI [31, 78]. Additionally, it is crucial to explore whether stakeholders can connect concrete system functionalities (such as selective disclosure, credential revocation, or wallet recovery) to the abstract qualities these systems promise [7, 20].

Accordingly, this thesis is guided by two research questions:

- **RQ1:** Which qualities are important for each category of users of DI and SSI systems?
- **RQ2:** Are the described functionalities of quality requirements clear to the users?

By answering these questions, the study seeks to bridge the gap between the theoretical ideals of SSI and the real-world expectations of its stakeholders, contributing to more aligned, trustworthy, and widely adoptable digital identity solutions.

1.2 Thesis Goals

This thesis is organized into six primary phases, each focused on a specific research objective.

1. **Background on DI and SSI:** Establish a robust foundation by exploring DI management systems, their architecture, and the processes involved in credential issuance, verification, and revocation. This examination will clarify the roles of identity holders (users), issuers, and verifiers, situating them within the trust triangle of SSI. The objective is to provide the necessary theoretical background to support subsequent phases.

2. **Background on Questionnaire Design:** Review the principles of effective survey and questionnaire construction in requirements engineering and related fields. This includes identifying existing approaches for measuring NFRs and adapting them to the DI/SSI context, with a focus on clear wording, neutral phrasing, and accessibility for diverse respondent groups.

3. **NFR Categorization:** Utilizing established quality frameworks for SSI (e.g., [98]), assign each NFR to one or more of the three system components: identity holder (user), issuer, and verifier. Each assignment should be substantiated by mapping responsibilities as primary, secondary, or tertiary, indicating the extent to which a role influences or depends on that requirement. Additionally, this process includes identifying representative entities for each role, such as citizens as users, universities as issuers, and employers as verifiers.

4. **Design and Operationalization of Questionnaires:** Create three role-specific surveys aimed at capturing stakeholder priorities regarding NFRs. Each questionnaire will ask participants to both (i) rate the importance of each quality and (ii) evaluate whether specific functionalities adequately fulfill that quality. To ensure clarity, abstract concepts will be transformed into concrete system functions (e.g., representing privacy through selective disclosure, and availability by allowing access to credentials at any time). The design of the surveys will rely on the Software Quality Requirements Importance (SQRI) scale, Best-Worst Scaling (BWS), and phrasing inspired by the Kano model, ensuring that the questions remain reliable, unbiased, and appropriate for multilingual contexts (English and German).

5. **Questionnaire Distribution and Data Collection:** Distribute the role-specific surveys to representative stakeholders using multiple channels (e.g., email, social media, professional mailing lists, direct outreach). The goal is to collect a diverse and balanced dataset covering all three SSI roles.

6. **Data Analysis and Discussion:** Analyze the collected responses using descriptive and inferential statistical methods. The analysis focuses on two research questions:
   RQ1: Which qualities are important for each category of users of DI and SSI systems?
   RQ2: Are the described functionalities of quality requirements clear to the users?
   Results will be visualized in a prioritization matrix that combines importance and

clarity measures, and will be compared to existing expert-based prioritizations [98]. Finally, construct validity and reliability will be assessed (e.g., Cronbach's α), and potential limitations in methodology or sampling will be critically discussed.

Together, these goals aim to contribute both empirical insights and methodological advances: providing a structured understanding of how different stakeholders in DI systems prioritize quality requirements, and testing whether survey-based approaches can capture these perspectives reliably.

1.3 Thesis Outline

The thesis proceeds in five tightly linked phases that turn a broad RE problem into an empirical prioritization of quality requirements in DI/SSI.

1. **Phase 1 – Foundations.** Establish the technical and stakeholder foundations: how credentials are issued, held, and verified; which roles act (identity holder (user), issuer, verifier); and why these systems should be judged through NFRs such as privacy, control, and security rather than only organizational goals. This sets the frame for a user-centric view of qualities rather than a solely business-driven one.

2. **Phase 2 – Mapping NFRs to roles.** Map the NFR set from prior work to the three roles, documenting a justification for each assignment and compiling concrete organizations to target later (e.g., typical issuers and verifiers). This produces the sampling frame and the role-specific rationale needed for measurement.

3. **Phase 3 – Questionnaire design.** Translate abstract qualities into answerable, role-specific survey items that are clear, unbiased, and, where helpful, localized (EN/DE). Items capture both perceived importance and the extent to which specific functionalities realize a quality, guided by DI/SSI design patterns; the verifier, for example, can rate the importance of "accessing credential-issuer information at any time".

4. **Phase 4 – Distribution and data collection.** Distribute the surveys through appropriate channels to the identified representatives and collect a diverse dataset covering the three SSI roles.

5. **Phase 5 – Analysis and reporting.** Analyze responses to reveal cross-role priorities, visualize results in a prioritization matrix, and report measurement quality (e.g., construct validity and internal reliability with Cronbach's α) alongside study limitations.

Together, these phases satisfy the project milestones on theoretical grounding, design, data collection, and evaluation, and explicitly surface the complexity of aligning questionnaire design, sampling, and statistical analysis in this domain.

Chapter 2

Fundamentals

2.1 Background

2.1.1 From Centralized to Decentralized Identity

Digital identity management is experiencing a significant shift from traditional centralized and federated models to decentralized frameworks [32, 64]. In conventional centralized identity systems, organizations or identity providers (IdPs) hold and control user data and credentials, which users must repeatedly share with various services [64, 77]. Such concentration creates single points of failure, increases the impact of breaches, and erodes user privacy and trust because large datasets become attractive targets for attacks [22, 32, 72, 76, 116]. Federated systems distribute responsibilities across domains but still rely on third party IdPs to manage and authenticate user identities, leaving users dependent on intermediaries [32, 93]. In contrast, DI systems place the individual at the core of identity management, minimizing reliance on central authorities and enabling users to control their own identity data [18, 93].

2.1.1.1 Principles of SSI

SSI empowers individuals by granting them ultimate authority and control over their digital identities [93, 97]. In an SSI model, identity is not granted or mediated by a central provider; instead, individuals accumulate verifiable claims about themselves such as their age, qualifications, or memberships and manage these claims in a personal digital wallet [93, 97]. SSI operationalizes DI through principles of user ownership and control, consent, minimal and selective disclosure, and interoperable, verifiable proofs across domains [97]. This user centric approach contrasts sharply with centralized systems where users have limited say in how their data is disseminated [93].

2.1.1.2 Core Components of Decentralized Identity Systems

DI systems rely on several core technical components. Decentralized Identifiers (DIDs) are globally unique identifiers that serve as an anchor for a user's identity without requiring any centralized registration authority [17, 64]. Unlike traditional identifiers such as email addresses or national ID numbers, which are dependent on centralized registries, DIDs remain under user control [64].

Verifiable credentials are the digital, cryptographically verifiable analogues of paper or plastic credentials people use in real life such as driver's licenses, passports, employee IDs, or diplomas [64, 97, 100]. A verifiable credential is cryptographically bound to the credential subject, typically a DID, and is signed by the issuer to certify the validity of the claims [64, 93]. Credentials contain metadata about their issuer, the subject, and validity period, enabling verifiers to assess trustworthiness [100].

The third vital component is the distributed ledger or decentralized registry that underpins the trust infrastructure for DIDs and credentials [18, 93]. The ledger, which could be a blockchain or another decentralized network, serves as a public key directory and as a trust anchor by publishing issuer DIDs, verification keys, credential schemas, and status information [17, 93, 97]. This public infrastructure enables verification without relying on a central authority [93].

Wallet applications hold private keys and credentials, orchestrate issuance and presentation, and implement secure agent-to-agent messaging [18, 77]. Wallets must support recovery and backup mechanisms so that users can regain access to their credentials if they lose device access, while maintaining security of cryptographic keys [18, 97].

2.1.1.3 Roles in DI/SSI

DI systems are characterized by a set of roles that different actors play in the ecosystem. The primary roles are the issuer, identity holder, and verifier, a triad sometimes depicted as a trust triangle [93, 97].

An issuer is an entity that creates and issues verifiable credentials to holders [93, 97]. This role is typically filled by organizations or authorities that have the legitimacy to attest to certain information about an individual, such as a university issuing a diploma, a government agency issuing a driver's license, or an employer issuing an employment confirmation [10, 97]. The issuer signs the credential with its private key, thereby vouching for the authenticity and integrity of the claims it contains [64, 97]. In DI, issuers do not need to maintain ongoing control over the credential or manage how it is used; they provide a trustworthy attestation at the time of issuance, and the credential, once issued, is under the control of the holder [97]. Issuers can later revoke the credential if it becomes invalid via the system's revocation mechanisms rather than by retrieving the credential from the user [64, 93].

The identity holder is the individual or entity who owns and controls verifiable credentials about themselves [64, 93, 97]. In many contexts, the holder is the user or subject of the

identity information [88, 93]. The holder's responsibility is to store credentials securely, typically in an identity wallet application, and to decide when and with whom to share those credentials [97]. The holder is at the center of all digital identity interactions in an SSI framework; nothing is shared without the holder's consent or initiation [88]. The holder can aggregate credentials from multiple issuers and present any subset of these as needed, enabling selective disclosure and minimal data exposure [31, 64]. The holder also safeguards their own private keys and recovery mechanisms; in SSI, if a user loses control of their keys, they effectively lose access to their identity data [93, 97].

A verifier is an entity that requests and verifies credentials presented by a holder in order to make an access or trust decision [93]. For instance, a verifier could be a security officer verifying a traveler's passport credentials, a website verifying a user's age before selling a restricted product, or a company's HR system verifying a job candidate's certifications [93]. The verifier specifies what information or proof is required and then cryptographically validates the received proof or credential [31, 64]. One key property of SSI is that the verifier does not need to contact the issuer directly to perform these checks because public keys, schemas, and status information are available via decentralized registries, reducing data leakage and coupling between parties [93, 97].

These three roles operate together to enable trustworthy identity transactions without a centralized intermediary [93, 97]. This model is often illustrated as a trust triangle: the issuer trusts the holder by granting them a credential, the holder trusts the issuer to provide a valid credential, and the verifier trusts both the issuer to have issued a valid credential and the holder to present their own credentials honestly [10, 93]. Trust is primarily handled through cryptography and decentralized infrastructure rather than institutional agreements [18, 93].

2.1.1.4 Core Processes in SSI

Credential Issuance
In the issuance phase, an issuer encodes claims according to an agreed credential schema, binds those claims to the holder's identifier via the credentialSubject.id field, and digitally signs the credential before delivering it to the holder's wallet [93, 97, 100]. Issuers can set validity constraints such as validFrom and validUntil, which verifiers later evaluate during checks [10, 100]. In practice, issuance is often mediated by offer and request flows and relies on publicly accessible schemas so that verifiers can later validate structure and authorization to issue [10, 26, 60].

Presentation and verification
When a verifier requests evidence, the holder's wallet composes a verifiable presentation containing selected claims or derived proofs such as selective disclosure or zero knowledge proofs and signs a challenge or nonce to demonstrate possession and holder binding [64, 90, 97]. The verifier then validates the issuer's signature and resolves the issuer's DID Document to obtain the correct verification keys [64, 90], checks holder binding and freshness of the presentation [2, 10], evaluates disclosed attributes or zero knowledge proofs [28, 64, 81], and consults a status registry to ensure the credential is not revoked or expired

[10, 64]. These checks can be performed without directly contacting the issuer because public keys, schemas, and status information are available via decentralized registries [93, 97, 100].

Revocation
To indicate that a credential is no longer valid, issuers publish status information that verifiers consult during policy evaluation [64, 97]. Widely referenced approaches include W3C Bitstring Status List v1.0 and the earlier Status List v2021 and Revocation List 2020 mechanisms, which aim to be privacy preserving, space efficient, and cacheable [99, 100]. For higher privacy, accumulator based schemes from the anonymous credentials literature enable revocation proofs without enumerating specific credentials [13]. Organizational guidance emphasizes deterministic handling of unknown status, distinguishing between soft fail and hard fail policies, and supports permanent or temporary revocation with reasons and history where appropriate [10]. Designs seek to minimize information disclosure about holders while ensuring universal detectability of revocation at verification time [2, 64, 100].

2.1.1.5 Authentication with DIDs

Beyond verifying claims, holders can authenticate by proving control of a DID, often called DID Auth [17, 93, 97]. A relying party issues a challenge; the wallet signs it with a key referenced in the DID Document; and the verifier resolves the DID to validate the signature and binding, often alongside or preceding a verifiable credential presentation [17, 31, 93].

Engineering studies show that caching DID Documents significantly reduces latency, while remote ledger lookups during handshakes increase costs [31]. These are practical considerations when deploying DID based authentication in protocols such as TLS [31]. For web scale interoperability, ecosystems are also adopting OpenID family protocols such as OpenID for Verifiable Presentations to standardize request and response flows between verifiers and wallets [11, 96].

2.1.2 Questionnaire Design Principles for Non Functional Requirements

Effective questionnaires are a fundamental tool for gathering stakeholder input in research, including requirements engineering [30, 49, 95]. Designing an instrument carefully is critical: a poorly constructed questionnaire can lead to biases and errors, whereas a well designed survey encourages respondents to provide accurate and adequate answers [95].

Clarity and simplicity in wording are central [30, 49]. Questions should be concise, focus on a single idea, avoid jargon and vague quantifiers, and steer clear of double negatives or leading phrasing [19, 95]. Practical guidance includes defining each construct up front, then iterating on item wording until a single interpretation is most likely [30, 49].

Minimizing bias is another crucial principle. Questionnaires must avoid leading or loaded questions and keep tone neutral [49, 95]. For construct validity, that is ensuring items actually measure the intended concept, authoritative guidance recommends explicit construct definitions, expert review, and evidence from pilot testing [39]. Precise response options, for example specific time frames rather than vague terms like "frequently," reduce ambiguity and improve interpretability [95].

The selection of response formats should follow your constructs and planned analyses. Closed ended items enable standardized analysis; open-ended items elicit richer qualitative detail [30, 95]. Likert type rating items are a staple for assessing attitudes and perceived importance. Evidence compares scale lengths: 5 or 7 point scales often balance respondent burden with reliability and discriminating power [80]. Including a neutral midpoint can help respondents express genuine neutrality when appropriate, though some respondents may misuse it as a default; it can be omitted if forced choice is necessary [80].

The overall structure and flow of a questionnaire significantly influence data quality. A brief introduction should explain purpose, confidentiality, and instructions, improving cooperation and response rates [19, 95]. A funnel approach, progressing from general to specific or sensitive topics, helps mitigate order and context effects [49, 95]. Clustering related questions with clear headings improves coherence; avoid question sequences that inadvertently cue or pressure answers [95].

Pretesting is essential to ensure survey validity. Conduct cognitive interviews to detect comprehension issues and pilot with a small, representative sample to refine wording, sequencing, and layout [79, 102]. For early instrument development, pilot sizes of approximately 15 to 40 participants can be adequate to identify potential issues with wording, formatting, or sequencing prior to full deployment [38, 45].

Surveys should be optimized for mobile devices and include explicit assurances of confidentiality, as these factors are known to influence response rates [63]. Recruitment strategies should be tailored to the intended sample through email lists, professional groups, and social networks, and plan pre-notifications, reminders, and incentives per the Tailored Design Method [19, 95]. For multilingual samples, translate carefully and verify through pretesting [79].

2.2 Related Work

2.2.1 Non-Functional Requirements in Requirements Engineering

NFRs in requirements engineering refer to software quality attributes, such as security, usability, and interoperability, that span different features and influence multiple stakeholders. These qualities are often challenging to specify and may conflict, making it essential to prioritize them early in the development process [36]. It is typically impossible to fulfill all NFRs to the same extent; trade-offs between qualities, such as security and usability, or among stakeholders' interests are necessary [36]. Structured questionnaires are practical tools for operationalizing and prioritizing NFRs by converting abstract

attributes into measurable statements that stakeholders can rate [82]. For example, [82] used Likert-scale survey items to quantify team members' views on the usefulness and ease of tools for managing NFRs, enabling the measurement of subjective preferences. These surveys clarify which qualities stakeholders consider most important and highlight areas of consensus or divergence [82]. By collecting stakeholder perceptions, questionnaires translate vague quality concepts into actionable data that guide engineering decisions. If users consistently prioritize privacy, development resources can be focused on solutions such as data minimization and encryption [82, 95]. Similarly, low usability scores indicate a need for better UI/UX design [95]. Overall, survey-based elicitation and prioritization of NFRs help requirements engineers make informed trade-offs and align system qualities with stakeholder values.

2.2.2 Measuring Quality Requirements: Existing Instruments

Although NFRs describe system qualities rather than user attitudes, many directly influence the user experience and can be assessed using established survey instruments [3, 42]. Utilizing these instruments offers two main benefits: they provide validated items for measuring specific qualities, and they allow for statistical testing of reliability and validity to assess how well the questions capture the targeted attribute. Below are some commonly used instruments for evaluating software quality:

Usability and Quality in Use are commonly assessed with standardized survey instruments. The System Usability Scale (SUS) is a widely recognized 10-item questionnaire for quick measurement of perceived ease of use [12]. The Post-Study System Usability Questionnaire (PSSUQ) evaluates user satisfaction across dimensions such as system usefulness, information quality, and interface quality [56]. To minimize respondent burden, shorter instruments like the Usability Metric for User Experience (UMUX) and UMUX-Lite maintain strong correlations with SUS scores while using fewer items [27]. The User Experience Questionnaire (UEQ) extends beyond usability by capturing both pragmatic factors (efficiency, dependability) and hedonic aspects (stimulation, novelty) [53]. Furthermore, models and standards such as QUIM and ISO/IEC 25010:2011 define usability sub-attributes and offer frameworks for linking these qualities to concrete survey items [1, 91].

Privacy, Trust, and Security Behavior are often measured using validated psychometric scales that align with NFRs such as data protection and trustworthiness. The Internet Users' Information Privacy Concerns (IUIPC) survey quantifies user concerns about data collection, control, and awareness [62], while the Concern for Information Privacy (CFIP) scale focuses on organizational data practices such as unauthorized access and improper use of personal information [92]. Trust in online services can be measured using scales that assess initial trust, disposition to trust, institution-based trust, and specific beliefs like competence and benevolence [65]. Security behaviors are captured by the Security Behavior Intentions Scale (SeBIS), which assesses habits such as password management, software updates, device protection, and attention to security indicators [23, 24]. These instruments can be adapted to the context of DI to assess user concerns about SSI wallet data protection or trust in issuers and verifiers.

2.2.3 NFRs in Decentralized Identity Systems

DI, especially in the form of SSI, transfers control over identity data from central authorities to individuals and the organizations directly involved with them. Standard SSI architectures rely on components such as Decentralized Identifiers (DIDs), verifiable credentials (VCs), user or enterprise wallets (agents), and a verifiable data registry, often a distributed ledger, to enable credential issuance, presentation, and verification without a centralized intermediary [64, 93]. The SSI ecosystem is structured around distinct roles: issuers create and vouch for credentials, holders manage credentials and consent to their sharing, and verifiers check credential authenticity and status, typically using the issuer's public DID or a revocation registry [64, 93].

This technical foundation is complemented by a set of NFRs central to SSI systems. Key attributes include *User Control* (ownership and autonomous management of identity data), *Privacy* and *Minimal Disclosure* (restricting shared information and protecting personal data), *Security* (preventing tampering and misuse), *Interoperability* (ensuring compatibility of credentials and DIDs across platforms), *Usability* (convenient identity interactions), and *Availability* (ensuring access to identity services when needed) [78, 98]. These NFRs reflect the core principles of SSI and are used as criteria for evaluating SSI solutions [78, 98]. Significantly, the perceived importance of these NFRs often varies by stakeholder: end-users are likely to prioritize *Usability* and *Privacy*, while verifiers focus on *Security* and *Reliability*. At the same time, issuers may emphasize *Interoperability* and compliance.

The following paper by [98] systematically identifies and classifies the quality attributes relevant to SSI systems. [98] developed a comprehensive taxonomy of SSI properties, including *Ownership and Control*, *Security* and *Protection*, *Privacy* and *Minimal Disclosure*, and *Interoperability*, which closely mirror classic NFR categories. Their taxonomy, validated via expert survey, serves as a robust reference for the qualities that SSI systems should ideally provide [98]. Crucially, this taxonomy forms the foundation of the present master's thesis, providing the complete set of 24 NFRs, which were mainly derived from this research by [98]. However, because the initial validation relied on expert opinion, the importance rankings may not reflect the perspectives of real end-users or organizations that interact with SSI systems in practice. This expert-centric, exhaustive taxonomy underscores the need for further investigation into stakeholder-specific priorities and perspectives on SSI NFRs.

Another line of related work focuses on design patterns in SSI architecture that operationalize quality attributes. In this context, design patterns are reusable solutions to common problems, translating abstract principles such as *Privacy* and *Recoverability* into concrete technical or procedural mechanisms. For example, [58] identified 12 key SSI design patterns that address challenges such as key management (including recovery and rotation), DID lifecycle (creation, update, deactivation), and credential presentation (e.g., selective disclosure). Building on this, [97] compiled a catalog of thirty-five SSI design patterns, systematically organized by core components, DIDs, verifiable credentials, wallets/agents, and verifiable data registries, and mapped to relevant stakeholder roles (identity holder (user), issuer, verifier).

These design pattern catalogs served as a foundation for this master's study, providing practical frameworks for operationalizing NFRs. For example, *Privacy* is supported through patterns such as selective disclosure and minimal-disclosure credential design, enabling holders to share only necessary information. *Security* and *Verifiability* are addressed through mechanisms such as revocation registries and status checks, which allow verifiers to confirm credential validity [58, 97]. Patterns for *Usability* and *User Control* include social recovery, guardian-backed key recovery to help holders regain access after key loss, and *Consent* receipt patterns to ensure holders can approve the sharing of their data [97]. By bridging high-level NFR concepts and low-level implementation strategies, these pattern catalogs underpinned this thesis's approach for connecting abstract quality attributes to concrete design solutions in SSI systems.

What remains lacking in the literature is a clear understanding of which qualities matter most to different stakeholders in real-world SSI deployments. While existing taxonomies and design pattern catalogs enumerate which qualities are important and how they can be technically achieved, they do not answer the more practical question of prioritization when trade-offs must be made. For example, technical constraints prevent achieving both perfect privacy and perfect usability simultaneously. In that case, it is unclear whether designers should prioritize *Privacy* or *Usability*, and the answer likely varies depending on whether the stakeholder is a identity holder (user), verifier, or issuer.

Moreover, most of this work, including the taxonomy by [98], focuses primarily on expert perspectives rather than end-users, leaving a gap in the literature regarding the actual priorities of individuals and organizations that interact with SSI systems. Early evidence from enterprise SSI pilots points to divergent stakeholder priorities. For instance, [10] found that organizations exploring SSI adoption emphasized governance frameworks, trust in credential issuers, integration with existing processes, and user training as key concerns. These priorities often result in trade-offs: an issuer may favor rigorous verification for security reasons, which can make the process more complex and less user-friendly, potentially reducing user adoption [10]. Similarly, verifiers may seek comprehensive credential data and broad interoperability, while holders prioritize minimal data sharing and robust user control for privacy.

Thus, beyond simply listing SSI NFRs, it is essential to understand how each role within the SSI trust triangle (issuer, holder, and verifier) ranks these qualities. This insight is critical for guiding system design, standards development, and deployment strategies that balance the needs and preferences of all parties. The present study addresses this gap by measuring the perceived importance of each NFR from the perspectives of identity holders (users), issuers, and verifiers, and by analyzing where these stakeholder views align or diverge.

2.2.4 Questionnaire-Based Approaches to Requirements Prioritization

Each technique has its own advantages and limitations. AHP offers rigorous, consistency-checked prioritization but does not scale well with large numbers of requirements due to the exponential number of pairwise comparisons required [3, 42]. Simpler approaches like MoSCoW or rank ordering are easier to implement but lack fine detail and do not indicate the degree of differences between priorities [47, 83].

Importantly, research shows that many traditional prioritization methods are aimed at project managers or developers, emphasizing functional requirements and relying heavily on expert judgment for quality attributes [3, 42]. This is particularly relevant for SSI, where expert-driven techniques, such as the expert survey used in [98], may fail to capture the proper priorities and nuances of end-users and other non-expert stakeholders.

Survey-based questionnaires are widely used to elicit stakeholder priorities for quality requirements. Likert-type importance ratings translate NFRs into concise statements that respondents can evaluate on an ordinal scale, and items are often tailored to specific stakeholder roles (e.g., identity holder (user), issuer, verifier) and informed by design patterns [3, 42, 95].

To enhance reliability, mirrored or paired items inspired by the Kano model are included [8, 83]. The Kano approach traditionally assesses both positive ("How do you feel if feature X is present?") and negative ("How do you feel if feature X is absent?") framing to classify the perceived necessity of quality attributes [42, 83]. In the survey design, for each quality, one item is phrased positively, and a paired item takes an opposing stance or describes a problem scenario.

Comparing responses to these positive and negative phrasings allows assessment of consistency. Stakeholders who rate *Privacy* as highly important would agree with the need for minimal disclosure and also rate total disclosure as highly problematic. This mirrored approach helps detect acquiescence bias and straight-line answering, thereby improving construct validity [70]. It also includes an attention check, as inconsistent responses to paired items may indicate inattentive or random responding.

However, solely relying on Likert importance ratings, even with improved item design, can result in skewed data. Participants often rate most qualities as important or very important, especially in domains such as SSI, where attributes such as *Security*, *Privacy*, and *Usability* are universally considered desirable. To address this and encourage more apparent discrimination among priorities, the survey includes a Best–Worst Scaling (BWS) exercise, also known as MaxDiff [54, 61].

BWS is a trade-off elicitation technique in which respondents see a small subset of items, for example, four or five NFRs at a time, and are asked to choose the most important and least important item from that set [61, 89]. This process is repeated with different subsets of items. Statistical analysis then produces a ranking and relative weight for each quality across all respondents [61, 89]. The technique has several advantages. It prevents respondents from rating all items equally highly. It produces interval or ratio-scaled preference measures. It works well even if there are many different items to consider [54, 89].

BWS has become a valuable tool in requirements engineering for prioritizing qualities or features without the cognitive burden of comparing every possible pair of items, which is required by methods like AHP. By combining BWS with direct Likert ratings, this study obtains two complementary sources of data. Absolute importance ratings use a familiar agree-disagree scale. Relative priority scores capture trade-offs and show the order of preference among qualities [3, 42, 61]. The BWS results help validate and refine the Likert results. If a respondent rates most NFRs as very important, BWS requires them to identify which are most and least important clearly.

Prior work in the SSI domain using questionnaires remains limited but offers instructive examples. [97] conducted an online questionnaire to validate their taxonomy of SSI properties, focusing on domain experts. Participants were recruited through professional networks and W3C working groups and asked to rate the relevance of each proposed quality property and to provide additional feedback. This expert-focused survey confirmed that qualities such as privacy, security, and interoperability are broadly considered important, providing strong vetting for the property list itself [97]. However, because the participant pool consisted mainly of SSI architects and researchers, the study did not capture the priorities of everyday users or organizations, nor did it compare or rank the relative importance of the qualities.

Gathering and reconciling stakeholder preferences for requirements is a well-established challenge in requirements engineering, with a variety of techniques available for prioritization [3, 42]. Classic methods include the Analytic Hierarchy Process (AHP), which uses pairwise comparison of requirements to derive weighted priorities; cumulative voting approaches such as the $100 test, where stakeholders allocate budgets to different requirements; the MoSCoW method, which classifies requirements as Must, Should, Could, and Won't have; and simple ranking or scoring [3, 47, 83].

Table 2.1: Comparison of Requirements Prioritization Techniques

Technique	What it does?	Strengths for my study	Key limitations	Decision
Software Quality Requirements Importance Scale — SQRI	• Self-report importance ratings • Mirrored +/− wording to reduce response bias • 5-point Likert scale	• Directly quantifies perceived importance across Users/Issuers/Verifiers	—	• Use ✓ • (primary importance measure)
Kano *[3, 8, 42, 83]*	• Asks functional (positive) vs. dysfunctional (negative) forms for the same attribute (Kano pair) • Classifies into categories	• Clear +/− phrasing makes abstract NFRs concrete • User-centric focus • Quick to prioritize requirements • Models satisfaction and dissatisfaction	• Not suitable for suggesting new features • Scalability issues	• ✓ Inspired for +/− phrasing • Not full Kano-classification

Technique	What it does?	Strengths for my study	Key limitations	Decision
BWS (MaxDiff) *[54, 61, 89]*	• Repeated small sets; pick most vs. least important • Choice experiment for prioritizing objects	• Efficient for larger sets of items • Forces discrimination among items • Avoids rating-scale biases	• Needs several tasks and careful explanation/design	• Use ✓ • to complement SQRI
AHP — Analytic Hierarchy Process *[3, 8, 42, 47, 83]*	• Breaks down complex decisions • Pairwise comparisons for prioritization	• Could provide reliable and efficient results • Fosters clear understanding	• Severe scalability issues $O(n^2)$ • Time-consuming • Complexity and difficulty of use	• × too heavy across many NFRs and 3 roles
Simple Ranking *[3, 47]*	• Orders requirements numerically	• Simplicity and ease of use	• Poor scalability • Lacks detail on relative differences • Potential for unreliable results	• × BWS already gives stronger trade-offs
$100 Test *[3, 8, 52, 83]*	• Point allocation by stakeholders • Prioritization by total score	• Easy to understand and apply • Fast execution • Can be accurate	• Poor scalability for large projects • Vulnerability to manipulation	• × BWS gives cleaner trade-offs at scale
MoSCoW (Must/Should/-Could/Won't) *[3, 47, 83]*	• Categorizes requirements into priority groups	• Ease of use; scalable • Fast setup • Consistent and low effort	• Lack of grading within categories • Does not provide a total ordering	• × BWS offers greater precision and bias reduction

2.2.4.1 Chosen techniques

From the review of prioritization methods from Table 2.1, two techniques were selected as the core of this study: SQRI as the primary measure and BWS as the complementary trade-off experiment. In addition, elements of the Kano model were integrated at the level of item wording. These techniques were chosen because they balance interpretability with methodological rigor, scale to the number of NFRs and stakeholder roles considered, and address common sources of bias identified in the literature.

- **Software Quality Requirements Importance Scale (SQRI).** The SQRI was established as the principal measure of importance for this study within the framework of this thesis. It is a self-report instrument that captures participants' perceived importance of quality requirements on a 5-point Likert scale. Items are phrased in plain language and can be expressed in both positive and negative forms to minimize response bias. SQRI directly quantifies perceived importance across the three stakeholder groups (identity holders (users, issuers, and verifiers), making it well-suited to a role-specific analysis of NFRs.
- **Kano model (wording inspiration only).** The Kano model asks respondents to evaluate functional (positive) and dysfunctional (negative) forms of the same attribute,

classifying requirements such as "must-be," "attractive," or "indifferent" [3, 8, 42, 83]. While this approach provides user-centric phrasing and helps make abstract NFRs more concrete, it has known scalability issues and is less suitable for large sets of requirements. For this reason, the model was not applied in full. Instead, its phrasing strategy was used as inspiration for mirrored positive/negative item wording in SQRI.

- **Best–Worst Scaling (BWS/MaxDiff).** BWS was chosen to complement SQRI by forcing respondents to discriminate between items in small repeated choice sets [54, 61, 89]. In each task, participants select the most important and least important requirement from a small subset, producing ratio-scaled data that avoids common rating-scale biases. BWS is efficient even with larger item sets and is therefore well-suited to prioritizing the broad set of NFRs in this study. While it requires careful explanation and design, its strengths in scalability and reliability outweigh these challenges.

2.2.4.2 Rejected techniques

Other well-established techniques were reviewed but ultimately rejected due to limitations in scalability, precision, or applicability to the study context. While methods such as AHP, simple ranking, the $100 test, and MoSCoW have proven useful in certain requirements engineering contexts, they were assessed as less suitable for handling the larger item sets and multi-role structure of this thesis. Their exclusion is based on a critical evaluation of trade-offs between methodological strengths and the practical constraints of the study design.

- **Analytic Hierarchy Process (AHP).** AHP is a structured decision-making technique based on pairwise comparisons that breaks down complex problems into a hierarchy of priorities [3, 8, 42, 47, 83]. Although AHP can produce reliable results and foster a clear understanding, its scalability issues are severe: the number of comparisons grows quadratically with the number of items. Given the large set of NFRs and three distinct roles in this study, AHP was deemed too resource-intensive and therefore unsuitable.
- **Simple ranking.** Simple ranking methods require respondents to order requirements numerically [3, 47]. While easy to understand and apply, this method performs poorly with large item sets, does not capture relative differences between items, and may yield unreliable results. Since BWS provides a stronger and more precise trade-off mechanism, simple ranking was rejected.
- **The $100 test.** In the $100 test, respondents allocate a fixed budget of points or currency units across requirements, with total scores indicating priorities [3, 8, 52, 83]. The method is fast, intuitive, and relatively accurate; however, it scales poorly for large projects and is vulnerable to manipulation. Because BWS provides more robust trade-offs and is less prone to bias, the $100 test was not selected.
- **MoSCoW.** The MoSCoW method categorizes requirements into four groups: must-have, should-have, could-have, and won't-have [3, 47, 83]. Its strengths are simplicity, scalability, and fast setup. However, MoSCoW lacks granularity, as it cannot

distinguish relative priority within categories, and it does not provide a complete ordering of requirements. For this reason, it was rejected in favor of BWS, which offers greater precision and mitigates bias.

In the broader context of RE, [16] surveyed software professionals to explore challenges in eliciting NFRs during design thinking workshops. Participants were carefully pre-screened for relevant experience, the survey was administered in their native language (Portuguese) to increase clarity, and recruitment was targeted via LinkedIn to ensure a suitable sample [16]. While [16]'s study did not directly focus on NFR ranking, it highlights best practices for questionnaire-based research, including adapting survey language to the target audience, defining technical terms, and using targeted recruitment strategies to improve data quality.

Requirements engineering offers several survey-based prioritization techniques (e.g., Likert-type ratings and Best–Worst Scaling) [3, 42]. However, within the SSI literature, existing questionnaires primarily validate taxonomies or surface concerns and do not apply role-comparative prioritization methods at scale across identity holders (users), issuers, and verifiers [10, 97, 98].

By surveying identity holders (users), issuers, and verifiers directly, with a carefully constructed instrument grounded in best practices from survey design and requirements prioritization research, this thesis offers the first comparative view of how different roles value non-functional qualities in DI systems. This knowledge can empower developers and policymakers to focus on the attributes that matter most to each community, making it easier to align SSI systems with user expectations and support wider adoption.

Chapter 3

Methods

3.1 Design

This chapter describes the research design used to prioritize NFRs for DI and SSI applications. The design adopts a role-aware perspective that distinguishes the three core roles in the SSI ecosystem: identity Holder (user), Issuer, and Verifier. A set of twenty-four NFRs (see Table 3.1), derived primarily from prior SSI property classifications, has been taken from the core reference by [98] and forms the basis of the study. In several cases, closely related properties were separated for analytical clarity (for example, *Security* versus *Protection*; *Verifiability* versus *Authenticity*). This approach was taken not only to improve analytical distinction, but also to ensure that each NFR is atomic, representing a single, distinct quality rather than a composite. By keeping requirements atomic, the prioritization process avoids ambiguity and makes stakeholder feedback and quantitative analysis more precise. The overarching purpose is to understand which qualities matter to which role and to what degree, and to use this understanding to guide a role-specific empirical instrument.

The design proceeds in two foundational steps, which together form the NFR Categorization process. First, each NFR is assigned to one of the three roles using a structured responsibility scheme with three levels:

- **Primary:** A core responsibility that is performed directly by the component and guarantees the fulfillment.
- **Secondary:** Reflects a supporting role, where a component facilitates performance of another's responsibility.
- **Tertiary:** Refers to indirect responsibility, where a component benefits from or relies on others to perform.

A brief, literature-grounded justification is recorded for every assignment. Second, representative entities for each role (*e.g.*, issuers (public agencies and universities) and holders

Table 3.1: NFRs for DI/SSI roles (identity holder (user), issuer, verifier). Definitions adapted from [98, 106].

Key	Quality	Short Definition (All Roles)
NFR1	Accessibility	Ability to access and retrieve identity data.
NFR2	Authenticity	Source of identity data is trustworthy and provable.
NFR3	Autonomy	Ability to manage identity independently of third parties.
NFR4	Availability	Identity data is available whenever needed.
NFR5	Compatibility	Identity data works across legacy and modern systems.
NFR6	Consent	Data use requires explicit consent from the actor.
NFR7	Control	Manage and control who accesses identity data.
NFR8	Cost	Minimal resources, effort, and financial costs.
NFR9	Decentralization	No core function relies on a central party.
NFR10	Existence	Identity exists independently of other services.
NFR11	Interoperability	Data works across platforms, wallets, and agents.
NFR12	Persistence	Data remains valid and accessible as long as needed.
NFR13	Portability	Ability to securely transfer identity data.
NFR14	Privacy	Share only the minimum necessary information.
NFR15	Protection	Data secured against misuse and threats.
NFR16	Recoverability	Ability to restore identity data after loss.
NFR17	Representation	Create/use multiple identities for different contexts.
NFR18	Security	Robust protection, transmission, and authentication.
NFR19	Single Source	Actor is the authoritative source for their identity.
NFR20	Standard	Credentials follow open standards (e.g. DIDs, VCs).
NFR21	Transparency	Information about data use is clear and available.
NFR22	Usability	Data is used efficiently and intuitively.
NFR23	User Experience	Identity management is simple and user-friendly.
NFR24	Verifiability	Identity claims can be verified and trusted.

(individuals or organizations) are identified to define the sampling frame for the empirical study. These steps provide a bridge from a literature-based NFR set to a concrete data collection plan.

To render abstract qualities measurable, NFRs are operationalized into concrete, observable functionalities that stakeholders can judge in realistic decision contexts. Role-specific questionnaires are subsequently specified with attention to clarity (single-idea items, avoidance of double-barrelled or leading wording), randomization of item order, and bilingual presentation where appropriate. The instrument combines simple importance judgments in plain language with a prioritization component designed to induce trade-offs, enabling ratio-scaled comparisons across a larger attribute set. This approach, which includes bias reduction through careful wording and mirrored framings where useful, reflects the study's goals and methodological guidance captured during instrument

development.

Finally, the analysis plan, pre-specified at design time, encompasses role-wise importance estimates, priority estimation from the trade-off tasks, construction of a prioritization matrix, and reliability and validity checks, all in line with the project goals.

3.1.1 NFR Categorization

The categorisation of NFRs by stakeholder role was conducted in two stages:
(1) - NFR-to-Role Mapping and **(2) - Representative Entity Identification**.

Stage (1) - NFR-to-Role Mapping
Each of the twenty-four NFRs was assigned to one or more of the core roles: Identity Holder (User), Issuer, Verifier. Together with a responsibility level, the following scheme was applied: *Primary* denotes a core responsibility directly performed by the role that was essential for fulfilling the quality; *Secondary* indicates a supporting responsibility that facilitates the primary function of another role; *Tertiary* refers to an indirect responsibility where the role benefits from, or depends on, other actors or the system to realise the quality. A short, literature-grounded justification was recorded for every assignment. When no single actor can guarantee quality (for example, due to decentralization or ecosystem-wide cost efficiency), ownership was attributed to the system/infrastructure layer and marked as tertiary for all roles. This process follows the design brief and provides traceability from textual sources to role assignments.

Worked examples (rest in A.1)

- **Consent (NFR6).** All three roles carry primary and direct responsibility, reflecting the centrality of informed consent in identity processes. The identity holder must deliberately grant and be able to withdraw consent for the use or sharing of personal data. Verifiers must obtain and respect consent for any collection, processing, or validation step they perform, and should request only the minimum data required for a single transaction. Issuers must also obtain explicit consent from subjects before issuing credentials that contain personal information, and limit the data to what is strictly necessary. This multi-role primary assignment aligns with SSI properties (e.g., privacy, minimal disclosure) and with consent requirements highlighted in surveys of SSI and DI.

- **Persistence (NFR12).** Persistence was assigned as a primary responsibility only to the identity holder. Identity data and identifiers should remain valid for as long as required by the identity holder, and cease when explicitly revoked or removed. The identity holder thus determines duration and continuity, while issuers and verifiers rely on persistent identifiers and credential status mechanisms without controlling persistence themselves; these latter roles are therefore not assigned primary responsibility. This interpretation aligns with SSI descriptions of long-lived identifiers and holder-centric control, as discussed in the literature.

Stage (2) - Representative Entity Identification
For each role, representative entities were catalogued to anchor the categorization in realistic contexts and to serve as the sampling frame for the empirical study. Typical issuers include public administrations, universities, certification bodies, and enterprises that issue credentials. Typical verifiers include employers, HR or compliance units, admissions offices, service providers, and public administration offices acting as relying parties. Holders include end-users (individuals) and, where applicable, organisations acting as identity subjects. The catalogue includes examples and use cases to inform recruitment and ensure that questionnaire items match operational realities.

Based on these role-based assignments, three separate surveys were developed, each tailored to reflect the role-specific functionalities of identity holders (users), issuers, and verifiers. Each NFR was translated into concrete, context-appropriate statements using a design pattern catalogues for SSI as bridges from abstract qualities to implementable behaviors. For example, to operationalize Privacy for issuers, the "selective content generation" pattern (minimizing data embedded in credentials) was applied to produce a concrete statement, such as: "Include only the personal data strictly necessary when issuing a credential." For verifiers, *Privacy* was reflected in requesting and accepting selective-disclosure proofs; for holders, *Privacy* became the ability to choose which attributes to present from the wallet. Using patterns in this way ensured that each responsibility assignment yielded a clear and realistic item that stakeholders could evaluate within their operational context.

All decisions (responsibility levels, ownership tags, justifications, and role-entity mappings) were maintained in structured spreadsheets to preserve end-to-end traceability from sources to survey items. These artefacts underpin the role-specific questionnaires and enable reproducibility of the categorization logic in later analysis and reporting.

3.1.2 Operationalization

In this phase, each abstract quality requirement was operationalized by translating it into specific survey items tailored for the three key stakeholder roles in SSI: Identity Holder (User), Verifier, and Issuer. This approach facilitated the connection of high-level NFRs, such as *Privacy* and *Security*, with tangible system functionalities. The objective was to ensure that each quality was easily comprehensible and evaluable for participants. By illustrating NFRs through straightforward examples or functional aspects, the questionnaire provided a consistent framework for interpretation among all respondents. Additionally, design patterns for SSI systems contributed to this process by linking abstract qualities to practical implementation features.

Notably, the catalogues of SSI design patterns found in the literature have been instrumental in mapping NFRs to functionalities A.2. [58] identified 12 patterns for blockchain-based SSI architectures, while [97] compiled a comprehensive collection of 35 SSI design patterns that encompass various facets of SSI ecosystems. These pattern libraries provide a knowledge base that informs the practical realization of specific quality attributes. In the adopted methodology, each NFR was linked with one or more pertinent design patterns (or architectural principles), ensuring that the survey items reflect realistic features that

effectively address those qualities. This approach aligns with the recommended practice of utilizing DI/SSI design patterns to clarify quality requirements.

For instance, the abstract concept of *Privacy*, defined as the principle of minimal disclosure of personal data, was translated into specific role-based functionalities. For Identity Holders, *Privacy* was operationalized as the capacity to selectively disclose information. The related questionnaire item emphasized sharing only the essential personal details necessary for identity verification. For Verifiers, *Privacy* was articulated as the ability to accept privacy-preserving proofs, indicating that verification should occur without accessing excessive personal data. This aligns with design patterns such as Selective Content Generation and Selective Content Disclosure [58, 97], which facilitate verification while revealing minimal information. For Issuers, *Privacy* was manifested as the issuance of credentials that prioritize data minimization and user consent, such as including only essential attributes and requiring the holder's approval for credential usage. Through these role-specific formulations, the abstract idea of privacy was transformed into concrete, observable behaviors and functionalities tailored to each stakeholder role.

The Table 3.2 showcases selected examples of how various NFRs have been translated into questionnaire items. Each entry comprises the NFR, the primary stakeholder role accountable for or affected by it, the design pattern or logic that influenced its transformation into functionality, and a paraphrased version of the related survey statement. These examples illustrate the diverse range of qualities being addressed, including *Security*, *Interoperability*, and *Usability*. They also demonstrate how theoretical quality requirements have been converted into practical, survey-ready constructs suitable for empirical evaluation. For full details of all operationalized NFRs and survey items by stakeholder group, see Table A.1, Table A.2, and Table A.3.

Table 3.2: Some examples of operationalized NFR items by role

NFR	**Role**	**Item Type**	**Design Pattern**	**Survey Item**
Privacy	Holder	Functional Importance	Credential Design Patterns - "Selective Content Generation" [58]; Verifiable Credentials and Presentations - "Selective Content Generation", "Selective Content Disclosure" [97]	"I want to share only the minimum details about myself when I prove my identity."
Privacy	Holder	Problem Importance	Credential Design Patterns - "Selective Content Generation" [58]; Verifiable Credentials and Presentations - "Selective Content Generation", "Selective Content Disclosure" [97]	"If I had to reveal more personal information than necessary, I would be fine."
Privacy	Verifier	Functional Importance	Credential Design Patterns - "Selective Content Generation" [58]; Verifiable Credentials and Presentations - "Selective Content Generation", "Selective Content Disclosure" [97]	"When I check someone's identity, I want to see only the data that are strictly needed for my service."
Privacy	Verifier	Problem Importance	Credential Design Patterns - "Selective Content Generation" [58]; Verifiable Credentials and Presentations – "Selective Content Generation", "Selective Content Disclosure" [97]	"Requesting more personal data than necessary would be acceptable for our service."
Privacy	Issuer	Functional Importance	Verifiable Credentials and Presentations - "Selective Content Generation" [97]; Credential Design Patterns - "Selective Content Generation" [58]	"I need to include only the personal data strictly necessary when issuing a credential."

NFR	Role	Item Type	Design Pattern	Survey Item
Privacy	Issuer	Problem Importance	Verifiable Credentials and Presentations - "Selective Content Generation" [97]; Credential Design Patterns - "Selective Content Generation" [58]	"Including any unnecessary personal data in a credential would be acceptable."
Security	Holder	Functional Importance	Key-Management Patterns - "Key Shards" [58]	"I need assurance that both the technology and its operators secure my identity data against threats."
Security	Holder	Problem Importance	Key-Management Patterns - "Key Shards" [58]	"I could handle the possibility of my identity data being exposed in a breach."
Security	Verifier	Functional Importance	Credential Design Patterns - "Blockchain Anchor" [58]; Trusted Registries – "Status Registry" [97]	"I need confidence that security threats to identity data are effectively managed during verification."
Security	Verifier	Problem Importance	Credential Design Patterns - "Blockchain Anchor" [58]; Trusted Registries - "Status Registry" [97]	"Effective security management during verification feels optional to me."
Security	Issuer	Functional Importance	Decentralised Identifiers & Cryptographic Keys - "Key Shards" [97]	"I must ensure that identity data are secure against threats during credential issuance."
Security	Issuer	Problem Importance	Decentralised Identifiers & Cryptographic Keys - "Key Shards" [97]	"Security measures during credential issuance feel nonessential to me."

The questionnaire's items were all based on the NFRs and followed the appropriate SSI design patterns or architectural principles. This alignment enabled people with different roles to associate each question with a specific function in a DI system, converting abstract qualities into measurable metrics. The operationalization step established a standard method for evaluating how well specific features, such as selective disclosure, audit logs, backup systems, and compliance with standards, meet the desired quality standards. By basing the items on well-established design principles, the survey was better able to determine the importance each NFR was perceived to have and the clarity of its implementation in the context of SSI. This operationalization was carried out thoroughly across the role-specific item sets: 24 NFRs for Identity Holders, 13 for Verifiers, and 12 for Issuers.

3.1.3 Requirements Prioritization Methodology

The study introduces the SQRI scale as a role-specific, plain-language instrument for assessing the perceived importance of NFRs in DI and SSI systems. As no existing scale was available for this purpose, SQRI was explicitly developed as part of this thesis to map each NFR to a concrete functionality per role and to function effectively in bilingual delivery. To reduce wording bias and ensure consistency in interpretation, a limited number of mirrored positive and negative stems were included [70]. This approach draws inspiration from the Kano tradition of functional versus dysfunctional phrasing, although no Kano classification was applied [3, 8, 42, 83]. To enable robust trade-offs and derive ratio-scaled priorities across larger item sets, SQRI was complemented by BWS (MaxDiff), an established choice-modelling method [54, 61, 89].

Methodology used in this study:
The survey instruments and measurement approaches are detailed in Section 3.3.

- **SQRI**: Items were formulated as short, single-idea, positively framed statements for each NFR and role. Simplified negative mirrors were applied only where clarity could be maintained. The item order was randomized.
- **BWS**: This trade-off measurement technique complemented SQRI by forcing stakeholders to choose the most and least important items from repeated small sets. The BWS design was tailored to each stakeholder group.

3.2 Participants

A total of 294 individuals started the survey across three stakeholder roles (identity holders (users), issuers, verifiers). After applying pre-registered cleaning rules, the final analytic sample comprised $N = 150$ valid participants (Identity Holders (Users) $N = 86$, Issuers $N = 37$, Verifiers $N = 27$).

The questionnaire had two main parts: Part 1 was the SQRI with FI and PI items. The second part was the BWS section. The SQRI was separated into two sections; the first included the NFR items, which were assigned a primary responsibility level, giving them higher priority. The second section included the NFR items with secondary and tertiary responsibility assignments. Survey participants with the `lastpage` ≥ 3 attribute set to 3 had completed the SQRI part but did not complete the BWS sections. Participants with `lastpage` $= 4$ or 5 completed the full survey.

Three exclusion criteria were applied:

1. **Incompleteness:** records with `lastpage` < 3 (i.e., Part SQRI not completed) were excluded.
2. **Straight-lining:** responses were excluded when the same Likert category was selected on all items within either SQRI block, corresponding to within-block response variance of zero.
3. **Duplicate submissions:** within each role file (identity holders (users), issuers, verifiers processed separately), submissions were identified as duplicates when response patterns were identical across all SQRI items and, where available, BWS choice fields; `seed/id` were also included when present. For each duplicate set, the first submission was retained and later repetitions were removed. This process was carried out manually, with the cleaned data set checked and duplicate entries removed.

Starting counts and exclusions were as follows:

- **Users:** initial $N = 146$; excluded $N = 60$ (incomplete = 58, straight-lining = 2, duplicates = 0); final $N = 86$. Of these, 80 completed both SQRI and BWS, and 8 completed SQRI only.
- **Issuers:** initial $N = 70$; excluded $N = 33$ (incomplete = 32, straight-lining = 1, duplicates = 0); final $N = 37$. Of these, 36 completed SQRI and BWS, and 2 completed SQRI only.
- **Verifiers:** initial $N = 78$; excluded $N = 51$ (incomplete = 40, straight-lining = 0, duplicates = 11); final $N = 27$. Of these, 24 completed SQRI and BWS, and 3 completed SQRI only.

3.2.1 Participant Characteristics by Role

Identity Holders (Users, $N = 86$)
From an initial pool of 146 submissions, $N = 60$ responses were excluded during data preparation ($N = 58$ due to incomplete questionnaires and $N = 2$ due to straight-lining; no duplicates were present). The analytic sample comprised $N = 86$ users, with $N = 80$ (93.02%) completing both survey sections (SQRI and BWS) and $N = 6$ (6.98%) completing only the SQRI. The mean survey duration was $M = 21.40$ minutes.

The mean age was $M = 31.60$ years ($SD = 13.20$), with ages ranging from 16 to 75 years. Gender distribution was: Female ($N = 42$, 48.84%), Male ($N = 39$, 45.35%), No Answer ($N = 5$, 5.81%); the "No Answer" option indicates participants who did not wish to disclose their gender, as the survey provided only "Male", "Female", and "No Answer" options. Regarding professional experience in DI or SSI, $N = 59$ (68.60%) reported no prior experience, while $N = 27$ (31.40%) reported relevant expertise.

Occupational categories were: In training (students/apprentices, $N = 43$, 50.00%), Professionals and Academics ($N = 17$, 19.77%), Clerical and Administrative ($N = 15$, 17.44%), Service and Sales ($N = 10$, 11.63%), Managers and Executives ($N = 9$, 10.47%), Technicians ($N = 5$, 5.81%), General Laborers ($N = 3$, 3.49%). As the job-category question allowed multiple selections, the summed percentages may exceed 100%.

Survey start language for the final sample was German ($N = 49$, 57.00%) and English ($N = 37$, 43.00%).

Verifiers ($N = 27$)
The verifier group tended from an initial pool of 78 submissions, $N = 51$ responses were excluded during data preparation ($N = 40$ due to incomplete questionnaires and $N = 11$ due to duplicate responses; no straight-lining was detected). The analytic sample included $N = 27$ verifiers, with $N = 24$ (88.89%) completing both survey sections (SQRI and BWS) and $N = 3$ (11.11%) completing only the SQRI. The mean survey duration was $M = 13.10$ minutes.

The mean age was $M = 44.90$ years ($SD = 10.50$), with ages ranging from 28 to 63 years. Gender distribution was: Male ($N = 15$, 55.56%), Female ($N = 9$, 33.33%), No Answer ($N = 3$, 11.11%). Professional experience with DI or SSI was reported by $N = 10$ (37.04%), while $N = 17$ (63.00%) reported no such experience.

Occupational categories included: Manager / Executive ($N = 21$, 77.78%), Professional / Academic ($N = 7$, 25.93%), Technician ($N = 2$, 7.41%), Clerical and Administrative ($N = 2$, 7.41%), Service and Sales ($N = 1$, 3.70%). As participants could select multiple job categories, percentage totals may exceed 100%.

Survey start language for the final sample was German ($N = 22$, 81.48%) and English ($N = 5$, 18.52%).

This group mainly included individuals from organizations involved in identity or credential verification. Many participants held positions in human resources or IT management. The verifier sample represented large enterprises and public institutions, as well as organizations from education, healthcare, and the private sector.

Common job titles among Verifier participants were in line with roles that handle credential verification processes:

- *Human Resources Management:* Head of HR, HR Managers/Consultants, Heads of Personnel responsible for verifying employee credentials or qualifications
- *IT and Security Management:* IT Managers, Heads of ICT, IT Architects (one specifically for IAM), IT Solution Architects, Information Security Managers who oversee systems that check digital credentials
- *Executive and Operations Management:* Managing Director/CEO, Head of Operations concerned with organizational identity verification processes
- *Process and Quality Control:* Process Manager, Quality Officer ensuring verification procedures meet standards
- *Training/Education Roles:* Training Manager (responsible for training/apprenticeships) verifying educational credentials of trainees

Issuers ($N = 37$)
From an initial pool of 70 submissions, $N = 33$ responses were excluded during data preparation ($N = 32$ due to incomplete questionnaires and $N = 1$ due to straight-lining; no duplicates were present). The analytic sample included $N = 37$ issuers, with $N = 36$ (97.30%) completing both sections of the survey (SQRI and BWS) and $N = 1$ (2.70%) completing only the SQRI. The mean survey duration was $M = 11.90$ minutes.

The mean age was $M = 46.80$ years ($SD = 11.30$), with ages ranging from 25 to 65 years. Gender distribution was: Male ($N = 31$, 83.78%), Female ($N = 3$, 8.11%), No Answer ($N = 3$, 8.11%). Prior professional experience in DI or SSI was reported by $N = 20$ (54.05%), with $N = 17$ (45.95%) reporting no such experience.

Occupational categories were: Manager / Executive (N = 23, 62.16%), Professional / Academic (N = 15, 40.54%), Technician (N = 6, 16.22%), Service and Sales (N = 2, 5.41%), In Training (N = 1, 2.70%). The job-category question allowed multiple selections, so category totals can exceed 100%.

Survey start language for the final sample was German (N = 27, 73.00%) and English (N = 10, 27.00%).

The issuer group consisted of professionals with significant experience in identity management and IT, many of whom held leadership roles. Participants worked in Swiss federal agencies, cantonal IT services, academic institutions, and private-sector companies and organizations involved in the identity and IT sectors.

Job titles among Issuer participants reflected high-level and specialized positions, for example:

- *IT and Security Leadership:* Chief Information Security Officers (CISOs), Chief Technology Officers (CTOs), IT Directors/Heads of IT (Leiter Informatik or Leiter IT), and a CEO
- *Identity Management Specialists:* Service Owners for Identity and Access Management (IAM), Enterprise Architects for large IAM systems, and e-ID domain specialists (e.g., "Fachspezialist e-ID" or Head of e-ID department)
- *Project and Product Managers:* Verifiable Credentials Project Coordinator, Head of IT Projects (Leiter Fachstelle Projekte), Product Manager, and team managers
- *Technical Roles:* Software Developers, Senior Software Engineers, Systems Administrators (including part-time roles), and IT Engineers
- *Security and Other Roles:* Security Architect, Security Officer

3.2.2 Recruitment and Procedure

Participants were recruited through targeted sampling strategies adapted to each stakeholder role. Issuer and Verifier participants were contacted primarily via email, often through the media or communication departments of relevant organizations. Recruitment focused on institutions and companies that either issue or verify digital credentials. These included Swiss federal authorities, cantonal and municipal IT services, higher education institutions, hospitals, and private-sector organizations across various domains.

The contacted entities covered a broad range of categories: universities, universities of applied sciences, and teacher-training colleges; small and medium-sized enterprises (SMEs) and technology companies; healthcare and pharmaceutical organizations; food and beverage producers; financial and insurance institutions; reinsurance and asset management firms; telecom and IT providers; engineering and industrial technology companies; luxury

and consumer technology brands; testing and certification services; logistics and transport IT firms; and government agencies at both federal and cantonal levels. Trust service providers and industry associations were also included.

Identity Holder participants (general users) were recruited through broader channels to reach a diverse population of digital identity users. Recruitment was conducted on social media platforms, including Instagram and Facebook, as well as through WhatsApp group chats and community forums. Additional responses were collected through workplace group chats at Hostpoint AG and on participant exchange platforms, including `SurveySwap.io` and `Poll-Pool.com`. SurveySwap and Poll-Pool were used exclusively to recruit Identity Holder participants. Both platforms are academic survey-exchange services that operate on reciprocal credit systems (SurveySwap uses "Karma" credits and Poll-Pool uses "PollCoins") to enable respondent exchange without monetary incentives. No payments or material rewards were offered in this study. Responses collected through these platforms were flagged and screened using the same inclusion and exclusion criteria as all other user entries. Across all roles, recruitment materials briefly introduced the study as a survey on digital identity. They invited eligible participants, such as those working in an issuer or verifier capacity, or users familiar with digital identity applications, to participate voluntarily.

3.3 Materials

All questionnaire items were derived from the comprehensive catalogue of NFRs compiled in this thesis, and were operationalized with reference to relevant SSI design patterns (e.g., credential minimization, status checking, key/backup management). Each item mapped an abstract quality requirement to a concrete, role-relevant functionality in the system. Two item types were used to measure the perceived importance of these qualities:

- **Functional Importance:** This item type assessed the perceived importance of a specific, tangible system feature. Each statement was phrased positively and described a desirable capability that implements the intended quality. Participants rated the importance of the described feature to them. Higher agreement indicated a higher perceived importance of the corresponding NFR.
- **Problem Importance:** This item type assessed respondents' concern about a negative outcome. Each statement was phrased negatively, describing a scenario where a functionality was missing or failed. Participants rated their agreement with the statement; lower agreement reflected greater concern about the potential problem and thus higher importance was attributed to the corresponding quality. Conversely, higher agreement indicated less concern about the absence or failure of that feature.

All items used a 5-point Likert scale with labeled anchors: 1 = Strongly Disagree and 5 = Strongly Agree, with 3 representing a neutral midpoint. Participants indicated their level of agreement with each statement. For the PI items, which were negatively worded, lower

agreement scores indicated greater perceived severity of the described problem and, therefore, higher importance attributed to the corresponding NFR. The consistent scale format, combining positively and negatively framed items, was designed to reduce acquiescence bias and other response biases.

This dual framing also functioned as a validity check of participants' understanding. For example, if a respondent rated access as very important on a positive item but also agreed that limited access is acceptable on the corresponding negative item, such an inconsistency would indicate potential misunderstanding or inattentive responding. Each statement was written in clear and straightforward language, focused on a single concept, and provided in both English and German to accommodate participants' language preferences.

Separate questionnaires were designed for each of the three key roles in the DI system: identity holder (user), verifier, and issuer. The content and number of items were tailored to each role's perspective:

- **Identity Holder Questionnaire (48 items):** The user questionnaire covered 24 NFRs from the holder perspective. Each NFR had a FI item. Where a clear adverse scenario could be formulated, a PI counterpart was added. Example (Protection, FI): "I need my identity data to be protected from unauthorized access or tampering during storage and transmission," informed by key management patterns such as Hot & Cold Wallet Storage [58, 97]. Example (Representation, PI): "I would be comfortable using a single digital identity in every situation," aligned with the Multiple Registration pattern for DIDs [58, 97]. A lower agreement indicates a higher importance of representation diversity. Example (Decentralization, FI): "I want to create and use new digital identities without depending on any single company or server," reflecting patterns such as Blockchain Anchor and Identifier Registry [58, 97]. The response scale ranged from 1 (Strongly disagree) to 5 (Strongly agree).

- **Verifier Questionnaire (26 items):** The verifier questionnaire included 26 items covering 13 NFRs relevant to verification activities. Each NFR was measured with two item types: a FI item and a PI item. For example, Authenticity was assessed with the positively framed statement "I must be able to prove that every credential I check is genuine and unaltered," which reflects the need for trustworthy validation and aligns with the Blockchain Anchor pattern [58, 97]. Compatibility was captured with the negatively framed statement "Rejecting a credential just because it was issued under another standard would be manageable for us," informed by DID management logic [58]. For PI items, lower agreement indicates a greater perceived severity and, therefore, a higher importance of the corresponding requirement. The response scale ranged from 1 (Strongly disagree) to 5 (Strongly agree).

- **Issuer Questionnaire (24 items):** The issuer questionnaire comprised 12 NFRs that corresponded to core issuer responsibilities. Each NFR was measured with two item types: a FI item and a PI item. For example, interoperability was assessed with the negatively phrased PI statement, "It would be manageable if some services rejected the credentials I issue," informed by the Trusted Schemas Registry pattern [97], which reflects sensitivity to cross-system compatibility. Protection was likewise captured with a negatively framed item, "A risk of interception or tampering of

credentials I issue is acceptable," aligned with key management patterns such as Hot & Cold Wallet Storage and Key Shards [58].

3.3.1 Best-Worst Scaling (BWS) Section

Following the SQRI items, each questionnaire included a Best–Worst Scaling (BWS/-MaxDiff) exercise. In each task, a small set of NFRs was presented, and participants chose the most and least important items. This forced-choice design elicited explicit trade-offs, reduced common rating-scale biases (for example, uniformly high ratings), and enabled ranking of the broader NFR set by relative importance. Item sets were randomized to limit order and context effects.

Each role had a tailored BWS design reflecting the number of NFRs relevant to that role:

- **Identity Holder:** The holder design included 24 NFRs and 18 choice tasks. Each task displayed four items. Participants selected the most important and the least important quality in each set. Every NFR appeared in exactly three tasks, enabling balanced exposure and estimation of relative importance scores for all 24 qualities.
- **Verifier:** The verifier design covered 13 NFRs and 13 choice tasks. Each task contained three items. Participants identified the most important and the least important qualities. Each NFR appeared three times across the section.
- **Issuer:** The issuer design comprised 12 NFRs and nine choice tasks. Each task displayed four items. Participants identified the most and least important qualities. Each NFR appeared in three tasks overall.

This BWS design, in which each quality appeared multiple times across tasks, enabled the calculation of stable importance scores for each NFR by role. Combined with the Likert-based SQRI ratings, it provided a comprehensive view of stakeholder priorities. The SQRI items yielded absolute importance ratings for each quality from each role's perspective. The BWS results produced a rank ordering by forcing comparative choices and estimating ratio-scaled importance shares across all tested NFRs. Together, these measures support the prioritization of the larger set of requirements for DI systems, mitigate individual scale-use biases, and enable a comparison of which qualities are most critical across roles.

3.4 Procedure

The survey was done online using the LimeSurvey platform, which was safely hosted on the University of Zurich's servers. The questionnaire was available in both English and German, allowing participants to choose the language they were most comfortable with, making it easier for them to understand. For each stakeholder role (Identity Holder (User), Verifier, and Issuer), there were two links, one in each language, allowing participants to choose the one they preferred.

Upon accessing the survey, participants were first presented with an introduction and information about consent. They were informed that participation was voluntary and anonymous. No personally identifying information, such as names or contact details, was collected to ensure privacy protection. When participants opened the survey link, they first encountered an informed-consent screen. This screen clearly stated the study's purpose, assured respondents of the anonymity of their responses, and explained that participation was voluntary. The consent text noted that participants could withdraw at any time. It also emphasized that the study conformed to the ethical principles of the University of Zurich (UZH) [107] and relevant Swiss research guidelines [110]. After reviewing this information, only participants who agreed to the terms proceeded to the questionnaire. Eligibility required a basic familiarity with digital identity interactions in the relevant role. Issuers were expected to have experience in issuing digital credentials, verifiers were expected to validate credentials, and users were expected to use digital identity credentials in everyday contexts.

The survey itself had four major sections. First, a background questionnaire was used to collect demographic and contextual data. This included questions about gender, year of birth, occupation, level of professional experience in DI/SSI, and (optionally for issuers and verifiers) job title or employer. These fields were primarily multiple-choice or short-answer (open text for organization), with clear instructions. All questions about employment (title and employer) were explicitly optional to avoid identifying individuals. Second, participants rated a series of SQRI Likert-scale items. These items were statements about various NFR criteria, and each was answered on a standard five-point Likert scale (e.g., from "strongly disagree" to "strongly agree"). The SQRI items were organized into two blocks based on whether each NFR related to the participant's primary or secondary/tertiary responsibilities. Both positively and negatively worded items were included to reduce response bias and ensure consistent interpretation. This dual framing also served as a validity check of participant understanding. For example, if a respondent rated access as very important on a positive item but also agreed that limited access is acceptable on a negative item, that inconsistency indicated misunderstanding or inattentive responding. Third, respondents completed a BWS prioritization task. In each BWS subtask, a subset of NFR criteria was displayed, and participants were asked to choose the most important ("best") and least important ("worst") item from the set. This process was repeated across multiple rounds with different combinations of items. Finally, the survey concluded with a thank-you screen that acknowledged participation and provided any necessary debriefing or contact information.

To reduce potential biases, the order of items was randomized for each participant. In practice, all SQRI statements and all BWS subtasks were shuffled in a counterbalanced way so that the sequence varied across respondents. This randomization was implemented to minimize order effects, as recommended in questionnaire design (randomizing question order is known to reduce context effects and systematic bias [9].

All collected data were stored anonymously. No personal identifiers (such as names, email addresses, or IP addresses) were recorded by the survey system. The (optional) employer and job title fields were explicitly set as non-mandatory and treated as sensitive: respondents could skip them, and they were used only to contextualize responses in aggregate, never to identify individuals. All procedures complied with the ethical research and data

protection guidelines of the University of Zurich. Data collection complied with the cantonal Data Protection Act (IDG) and the University of Zurich Policy on Ethical Review of Research Projects Involving Human Subjects [114]. Participation was entirely voluntary, and respondents could withdraw or discard their responses at any point before submission. The information collected was used solely to examine how perceptions of NFRs varied across participant groups.

The survey system recorded the average completion times: Identity Holders took approximately 21.4 minutes, Verifiers took about 13.1 minutes, and Issuers took about 11.9 minutes to complete their respective surveys. These times reflect the longer questionnaire for Identity Holders. No formal blinding was used, as each participant knowingly completed the survey designed for their role; therefore, the role context was not concealed. Throughout the procedure, care was taken to ensure that each participant experienced the study consistently and ethically. Informed consent was obtained prior to data collection. The survey experience (including instructions, item phrasing, and timing) was standardized for all respondents, and all protocols conformed to the University of Zurich's ethical standards [114].

The overall procedure was designed to ensure a clear, consistent, and ethically compliant participant experience across all stakeholder roles.

Chapter 4

Results

4.1 Data Preparation

This study applied a rigorous data preparation and quality control process for survey responses, guided by domain practices [33, 40, 66]. Incomplete and clearly invalid submissions were excluded using a comprehensive cleaning strategy that also addressed straight-lining (identical answers across items), duplicate submissions, and internal inconsistencies between survey sections. The reproducible data cleaning pipeline centered on three pillars: exclusion of invalid cases, principled handling of missing data, and precise documentation of all transformations.

The survey design included two SQRI blocks, each featuring core NFRs in straightforward language and assessed with a five-point Likert scale. Positive and negative phrasings, purposefully mixed to mitigate acquiescence bias, were inspired by Kano's method. An additional BWS task presented 3–4 NFRs per screen to elicit trade-offs and generate interval-scaled importance scores (Users: 18 tasks; Issuers: 9; Verifiers: 13). These measures maximize reliability and provide a robust evidentiary basis for subsequent analyses.

Exclusion Criteria: Detailed exclusion procedures were applied immediately following initial participant counts. In keeping with established survey quality standards, three main categories of problematic records were removed: (1) incomplete questionnaires flagged by last-page progress (`lastpage` not in 3, 4, 5); (2) straight-lining responses with zero variance across Likert items within either SQRI block; and (3) duplicate submissions, identified by identical normalized response patterns, where only the first entry from a duplicate set was retained. Each removed entry is documented in an audit file with the apparent drop reason (see 3.2), ensuring full traceability.

Item-level missingness was minimized through the survey design, which required a response to every SQRI item and enforced choices for both "most important" and "least important" options on every BWS screen. This approach largely prevented item nonresponse. Any remaining missing data was almost entirely due to unit nonresponse, which was handled through last-page exclusion. As a result, the analytic data set is based on responses that pass all exclusion criteria, with sample sizes reported transparently for

each stakeholder group. Analyses were conducted using listwise deletion, consistent with best practices.

Observed exclusions and final analytic N.

Table 4.1: Participant exclusions and resulting analytic N per role.

Role	Incomplete	Straight-liners	Duplicates	Final N
Identity Holders	58	2	0	**N = 86**
Issuers	32	1	0	**N = 37**
Verifiers	40	0	11	**N = 27**

Overall, these exclusions yielded a final analytical sample of 150 respondents across all three stakeholder groups, ensuring that only high-fidelity data contributed to subsequent analyses. This rigorous data refinement process is crucial for enhancing the study's internal validity and ensuring the reliability of the insights derived from NFRs.

Consistency Checks: SQRI incorporated both positively and negatively framed items for selected NFRs, which served as an internal validity check. For example, if a respondent rated "access" as highly important while also agreeing that "limited access is acceptable," this indicated an internal inconsistency. Such inconsistencies were inspected during the screening process. When the pattern suggested inattentive responding, such as uniform or contradictory answers across the FI and PI blocks, the corresponding responses were classified as straight-lining and excluded according to the established data cleaning rules. Minor inconsistencies that did not indicate systematic inattention were retained to avoid unnecessary data loss.

Transformations and Derived Variables: To ensure comparability between positive and negative phrasings, the PI items are reverse-coded using the formula `PI_rev = 6 - PI`. This adjustment aligns both blocks to a consistent interpretation where "higher" indicates greater importance or impact. Prior to analysis, all Likert scale labels are converted to integers ranging from 1 to 5. Demographic information is standardized as follows: gender is derived from `Dem01` (represented by the initial uppercase letter), and `birth_year` is extracted from `Dem02`, accommodating standard date formats. Age is calculated as the current year (defaulting to 2025) minus `birth_year`. Any administrative and transient fields are excluded, specifically all columns containing "Time" along with `startdate`, `datestamp`, `submitdate`, and `ThankYou`. Identifiers such as `id` and `seed` are omitted by default unless explicitly retained. Consequently, the cleaned datasets consist of all FI/PI items, including the computed `PI_rev`, any BWS responses, standardized demographics (`gender`, `birth_year`, `age`, `startlanguage`), contextual variables necessary for subsequent subgroup analyses (e.g., `Erfahrung` (SSI experience), job indicators `Dem03*`), and the provenance flag `lastpage`.

FI and reverse-coded PI (PI_{rev}) are analyzed as separate dimensions rather than combined into a single score. FI measures how much participants value a specific quality being present, while PI_{rev} shows how much they are concerned about its absence or failure. Keeping these measures separate makes it possible to see cases where a quality is important but may not be delivered (High FI, Low PI_{rev}), or where it is well covered but not seen as critical (Low FI, High PI_{rev}). This separation gives a clearer and more nuanced basis for

prioritization, following widely used logic in importance–performance analysis for quality assessment.

Role-specific outputs and audit trail. The cleaning script is executed separately for each stakeholder role (identity holder (user), issuer, verifier). It produces paired artifacts per role: a lean, analysis-ready CSV file (*_cleaned.csv*) and a companion audit CSV file (*_audit.csv*)[1]. The audit file lists every excluded record along with its corresponding reason. These outputs feed an automated participant summary report that tallies initial counts, exclusion types, final valid cases, and descriptive demographics for each role (Identity Holders (Users) $N = 86$, Issuers $N = 37$, Verifiers $N = 27$). This reporting step reads the cleaned and audited files, recalculates completion status from the raw `lastpage` variable, and summarizes age, gender, survey duration, experience, and job categories.

[1]For access to the processed datasets and audit files, see Appendix B.

4.2 Analysis Overview

The quantitative survey data were analyzed through three key components, each focusing on a distinct aspect of the results. The Python script for SQRI processes Likert-scale responses, calculating means for each item and role, confidence intervals, reliability indices, and accompanying summary plots. The Python script for BWS estimates ratio-scaled importance shares from the BWS tasks using an MNL-based exploded-logit (rank-ordered) model, yielding rank orders, importance share tables, and comparative graphics for the three stakeholder groups. The Python pipeline for role comparison integrates these outputs to differentiate stakeholder perspectives, compile the prioritization matrix, and summarize cross-role disparities. This section details the analytical procedures, their rationale, the required data inputs, and the key tables and figures that support the Results chapter. In relation to the research questions introduced in Section 1.1, RQ1 is addressed primarily by the SQRI Analysis (Section 4.2.2), the BWS Analysis (Section 4.2.3), and the Cross-Role Analysis (Section 4.2.4), which together derive and compare importance profiles for identity holders, issuers, and verifiers. RQ2 is addressed by the internal reliability and construct checks (Section 4.2.1) and by subsequent analyses that relate functionality ratings to NFRs and compare prioritization patterns across methods and roles, particularly in the SQRI Analysis, BWS Analysis, and the Cross-Role Analysis.

Statistical analysis followed the procedures described on the University of Zurich's Methodenberatung website [111], which served as the primary methodological reference. Its guidance informed variable preparation, the selection of appropriate tests, checks of underlying assumptions, the reporting of effect sizes and confidence intervals, and the clear presentation of results.

4.2.1 Statistical Assumptions and Test Selection

The choice of statistical tests in this study is based on the type of survey data and careful checks of the assumptions each test requires. Because Likert-scale responses are ordinal and may not meet the assumptions of parametric tests (such as normality or equal variance), only nonparametric methods were used.

Before running these primary analyses, basic assumption checks were performed on each sample. For one-sample tests, the Wilcoxon signed-rank test [115] was chosen to compare median ratings to the neutral scale point (3 on a 1–5 scale). To check whether the Wilcoxon test could be trusted, the symmetry of responses around the midpoint was checked for each item. When this symmetry was weak, a Sign test was reported as a backup, since it does not rely on symmetry.

To compare groups, Kruskal-Wallis tests [109] were used both within and between stakeholder roles: within each role to examine differences by gender, professional role, and SSI experience, and across roles (identity holders, issuers, and verifiers) to test whether their NFR ratings differed systematically. This test is suitable for ordinal data and independent groups. When group sizes were small or group responses showed different shapes or spreads, results were interpreted carefully, focusing on differences in distributional shapes

and spreads. When a group difference was significant, post-hoc Dunn-Bonferroni [109] tests were used for pairwise comparisons.

To compare NFRs within respondents, the Friedman [108] test was used as a non-parametric repeated measures procedure on the FI and PI_{rev} block ratings, testing whether certain NFRs were systematically ranked higher or lower than others within the same stakeholder group. The analysis first ensured that enough respondents answered all relevant questions, ensuring the test would be valid. Kendall's W was reported as a measure of agreement for Friedman tests, when significant, Dunn–Bonferroni post hoc tests identify which NFR pairs differ.

Correlations between variables were analyzed using Spearman's rank correlation coefficient (Spearman's ρ) [113]. This method is well-suited for ordinal data and monotonic relationships and does not require assumptions of linearity or normality; however, the variables should be at least ordinal-scaled. Spearman's ρ enables the analysis of associations when variables are ranked or relationships are not strictly linear.

Other checks included examining whether many responses clustered at the highest or lowest scale values (ceiling or floor effects), which can make the means less meaningful. Because of these patterns, medians and interquartile ranges were most often reported. Effect sizes reported include r for Wilcoxon, ε^2_H for Kruskal–Wallis, Kendall's W for Friedman, and Spearman's ρ.

All these checks were performed and saved in structured Excel reports [2], so that decisions about statistical methods were justified and reproducible. The approach used here follows guidance for ordinal survey data and helps ensure that results are reliable and robust, given the data's actual properties.

Results of assumption checks

From the assumption checks, the identity holder (user) sample showed several items with right-skewed distributions and moderate ceiling effects, indicating that many participants selected high importance ratings. This pattern is typical for Likert data and supports the use of nonparametric tests. Group comparisons showed broadly similar distribution shapes across roles, confirming that the Kruskal-Wallis test was appropriate. For the issuer sample, responses were also skewed toward agreement, but the smaller group size led to greater variability in the distribution; results from this group were therefore interpreted with caution. For the verifier sample, the number of responses was lower, and distributions were again skewed toward high importance values. All Friedman tests had sufficient complete cases per role, so within-respondent comparisons of NFRs were valid. Overall, these checks confirmed that the data met the basic requirements for using Wilcoxon, Kruskal-Wallis, Friedman, and Spearman tests, and that nonparametric methods were the most reliable choice for this dataset.

All statistical procedures followed the guidance provided by the University of Zurich's Methodenberatung (engl. Statistical Consulting) resource [111], ensuring that test assumptions were appropriately verified and that results were robust to violations of parametric assumptions.

[2] For access to the Excel files, see Appendix B.

4.2.1.1 Internal reliability and construct checks - Identity Holders (Users)

Cronbach's α for FI and PI_{rev} (with α-if-deleted)
For identity holders (users), both the FI and PI_{rev} scales exhibited high internal consistency. Cronbach's α was 0.890 for the FI scale and 0.889 for the PI_{rev} scale, indicating good reliability in each case. Furthermore, removing any single item did not substantially affect these values: for both scales, α-if-deleted remained within a narrow range of 0.88 to 0.90. In fact, the item corresponding to *Representation* (NFR17) was the only one whose removal slightly raised α (to about 0.897 in both scales), suggesting it was somewhat less in sync with the others; however, even this improvement was minimal. Thus, all items were retained as each contributed meaningfully to overall consistency.

Table 4.2: Block-level reliability estimates (Cronbach's α) and α-if-deleted coefficients for the 24-item SQRI scales.

Block	Item / Statistic	α-if-deleted	Change (Δ)
FI	**Overall Block Reliability (24 items)**	**0.8898**	—
	Highest consistency increase if deleted:		
	NFR17 – Representation	0.8971	+0.0072
	NFR24 – Verifiability		+0.0049
	NFR11 – Interoperability		+0.0013
	Highest consistency decrease if deleted:		
	NFR13 – Portability		−0.0112
	NFR5 – Compatibility		−0.0111
	NFR8 – Cost		−0.0098
PI_{rev}	**Overall Block Reliability (24 items)**	**0.8894**	—
	Highest consistency increase if deleted:		
	NFR17 – Representation	0.8971	+0.0077
	NFR10 – Existence		+0.0004
	NFR9 – Decentralization		+0.0001
	Highest consistency decrease if deleted:		
	NFR20 – Standard		−0.0099
	NFR4 – Availability		−0.0086
	NFR5 – Compatibility		−0.0083

Note. The table reports the overall reliability for both blocks (FI and PI_{rev}) and highlights items with the highest and lowest changes in α if deleted. Positive Δ values indicate that removing the item increases internal consistency; negative values indicate a decrease.

The α-if-deleted diagnostics reveal that the removal of *Representation* (NFR17) would yield the most substantial reliability gain for both blocks, suggesting that this item contributes the least to internal coherence within the scale relative to the other items.

Only minor positive changes are observed for a few additional items, including *Verifiability* (NFR24) and *Interoperability* (NFR11) in the FI block, as well as *Existence* (NFR10)

and *Decentralization* (NFR9) in the PI_{rev} block, which indicates that potential improvements from the removal of these items are marginal when compared to the implications of removing NFR17.

The majority of items exhibit a reduction in reliability if deleted, with the most significant decreases associated with core attributes such as *Portability* (NFR13), *Compatibility* (NFR5), and *Cost* (NFR8) in the FI block, alongside *Standard* (NFR20), *Availability* (NFR4), *Compatibility* (NFR5), *Transparency* (NFR21), and *Security* (NFR18) in the PI_{rev} block. These findings highlight the integral nature of these items to their respective constructs.

In conclusion, the evidence supports retaining the complete item sets for both blocks, while also designating *Representation* (NFR17) as a candidate for further conceptual review rather than immediate exclusion based solely on reliability metrics.

Inter-item structure: Spearman correlations
Pairwise Spearman rank correlations (ρ) were computed for all SQRI items within each block (FI and PI_{rev}) to examine the internal structure and inter-relationships among the 24 NFRs. Spearman's rank correlation was chosen over Pearson's because the SQRI items use ordinal Likert-scale data. Spearman's ρ measures the strength and direction of monotonic relationships without assuming interval-level scale, normal distribution, or linearity, making Spearman's method well-suited for ordinal survey data [112, 113]. In contrast, Pearson's correlation is appropriate only for continuous, normally distributed variables. All reporting is based on Spearman's correlation, ensuring compatibility with the data's properties and robust results [112, 113].

With 24 items per block, the analysis yielded 276 unique item pairs per correlation matrix. To control for false positives, the Holm procedure was used to adjust significance thresholds across all comparisons. The full results, including correlation values, raw and adjusted p-values, and significance markings, are available in the project repository (see B). Figures 4.1 and 4.2 show heatmaps with stars for significant pairs, summarizing both the strength and reliability of the correlations.

FI block correlations
For the FI block, correlations were predominantly positive across the 24 quality requirement items, consistent with a general factor of perceived importance. The average off diagonal correlation coefficient was $\bar{\rho} = 0.27$, and 66 of the 276 item pairs (23.9%) remained statistically significant after adjustment at the .05 level. The strongest observed associations emerged among conceptually related requirements. The highest correlation was observed between NFR5 (*Compatibility*) and NFR13 (*Portability*), with $\rho = 0.637$, $p_{\text{adj}} < .001$, followed closely by NFR5 (*Compatibility*) and NFR22 (*Usability*) at $\rho = 0.628$, and NFR2 (*Authenticity*) and NFR13 (*Portability*) at $\rho = 0.611$, all of which were significant after adjustment. These high positive correlations reflect logical clustering: *Compatibility*, *Portability*, and *Usability* capture overlapping user facing technical requirements, while *Authenticity* and *Portability* both relate to credential integrity and transferability. A small number of item pairs exhibited weak negative correlations. The most negative

association was between NFR14 (*Privacy*) and NFR24 (*Verifiability*), with a correlation coefficient of $\rho = -0.173$. However, this correlation did not survive the adjustment ($p_{\text{adj}} > 0.05$). Such divergent ratings suggest that some respondents may perceive specific quality attributes, particularly *Privacy* versus *Verifiability*, as trade-offs or distinct priorities rather than as uniformly essential features.

Using conventional benchmarks for rank correlations (absolute ρ of .10 equals small, of .30 equals moderate, and of .50 or more equals large), the observed effects in the FI block span from trivial to large. The mean off diagonal correlation $\bar{\rho} = .27$ indicates small to moderate typical associations. The strongest pair, *Compatibility* and *Portability* (NFR5 and NFR13), exhibits a large association with $\rho = 0.64$, indicating a significant correlation. The weak negative pair *Privacy* and *Verifiability* (NFR14 and NFR24) at $\rho = -0.17$ is trivially small in magnitude. Among the 66 statistically significant pairs after adjustment, most lie in the .20 to .40 range, that is, small to moderate effects.

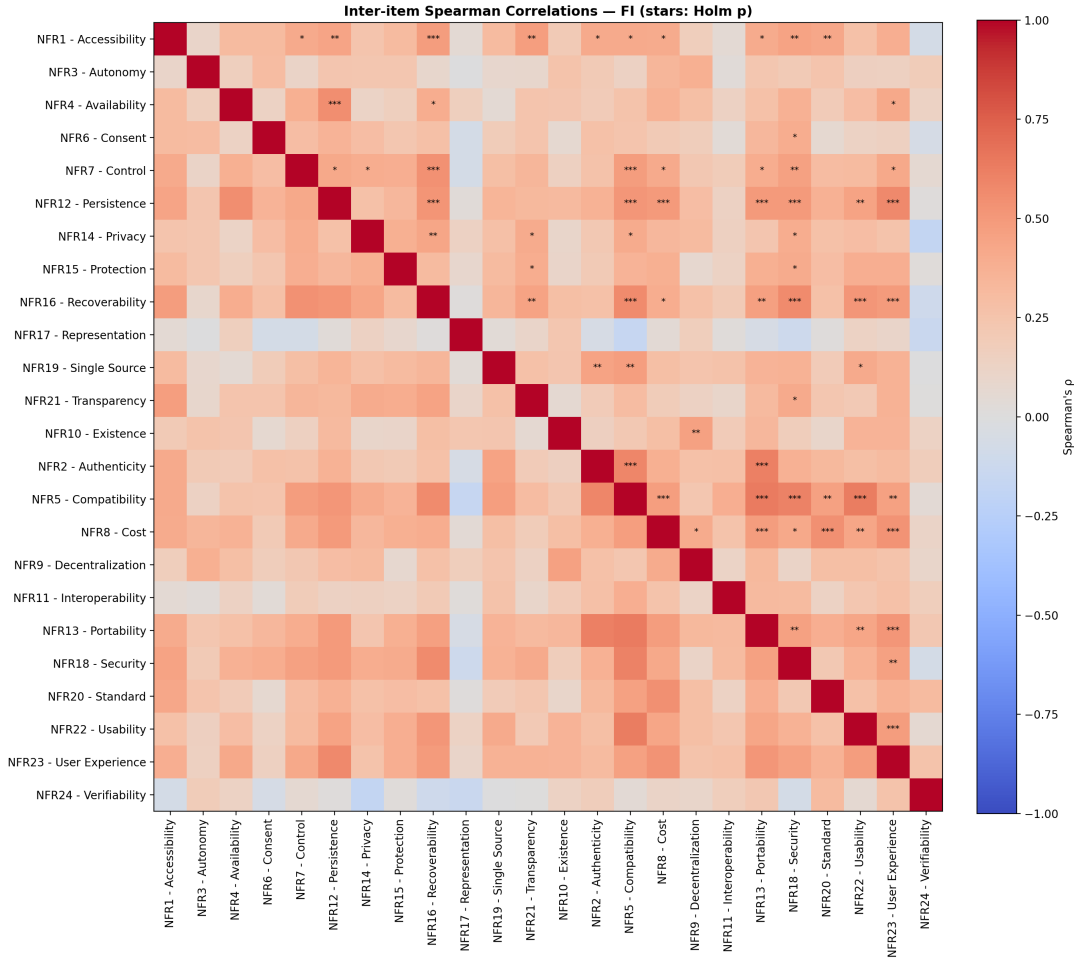

Figure 4.1: Spearman correlation heatmap for the FI block, showing ρ values with Holm-adjusted significance levels.

PI_{rev} block correlations
The PI_{rev} block displayed a similar overall pattern but with slightly lower average correlation strength. The mean off diagonal coefficient was $\bar{\rho} = 0.24$, and 40 of the 276 pairs (14.5%) were statistically reliable after adjustment. The strongest association was observed between NFR5 (*Compatibility*) and NFR11 (*Interoperability*), yielding $\rho = 0.721$, $p_{\text{adj}} < .001$, the highest correlation observed across both blocks. This robust positive relationship reflects respondents' perception that current DI solutions either deliver both technical integration capabilities effectively or fall short on both dimensions together. The following strongest correlations in the PI_{rev} block were NFR5 (*Compatibility*) with NFR20 (*Standard*) ($\rho = 0.563$) and NFR5 (*Compatibility*) with NFR13 (*Portability*) ($\rho = 0.538$), both significant after adjustment, reinforcing the interpretation that *Compatibility*, *Standard*, *Portability*, and *Interoperability* form a tightly coupled cluster in users' evaluations of implementation quality.

A subset of negative associations was present in the PI_{rev} block, more pronounced than in FI. The most negative correlation was between NFR17 (*Representation*) and NFR19 (*Single Source*), with $\rho = -0.304$, which did not reach adjusted significance but suggests potential divergence in how these lower priority items are perceived in practice. These negative associations were infrequent and often involved NFR17 (*Representation*), consistent with its marginal fit in the reliability analysis.

Applying the same benchmarks to the PI_{rev} block, the mean off diagonal correlation $\bar{\rho} = .24$ is small on average. The strongest pair, *Compatibility* and *Interoperability* (NFR5 and NFR11), has a correlation coefficient of $\rho = .72$, indicating a large effect. The negative pair *Representation* and *Single Source* (NFR17 and NFR19) at $\rho = -0.30$ is moderate in absolute magnitude. Of the 40 significant pairs after adjustment, the majority again fall in the small to moderate band.

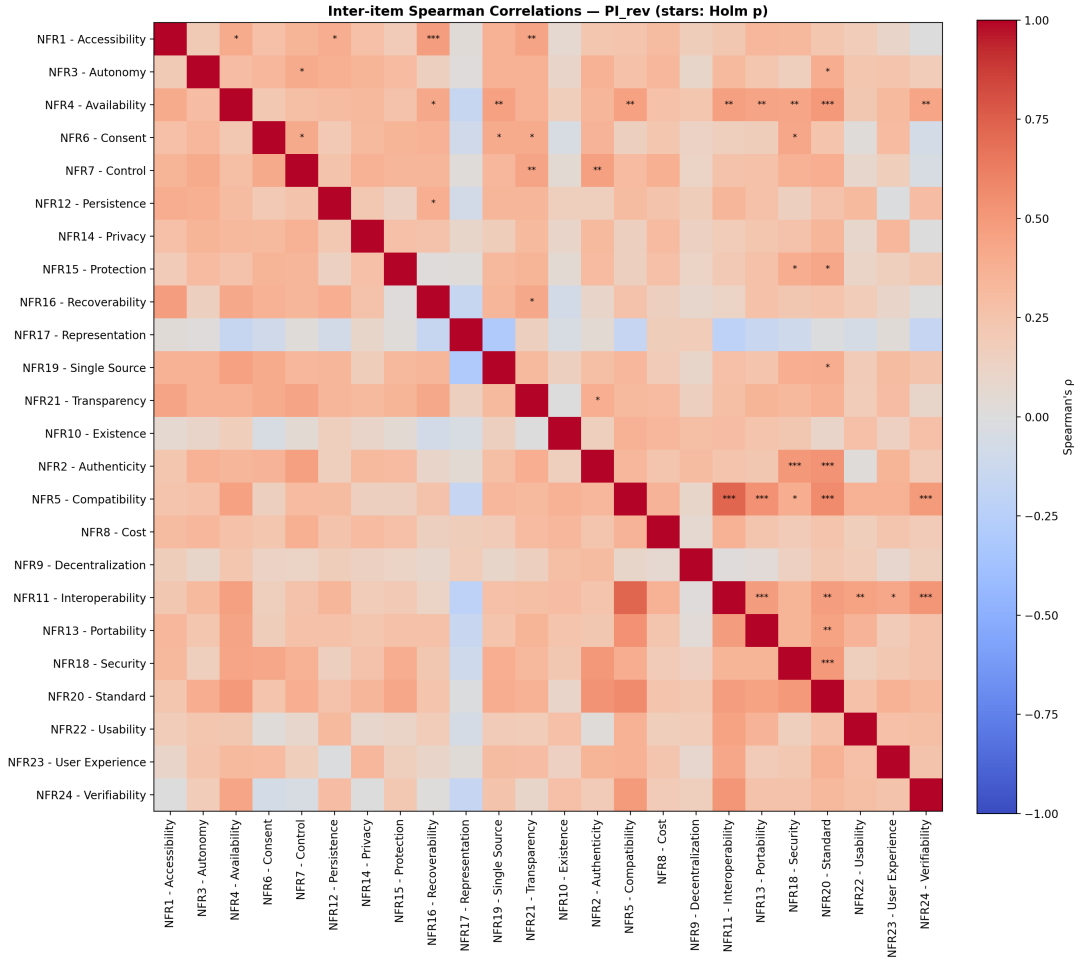

Figure 4.2: Spearman correlation heatmap for the PI_{rev} block, showing ρ values with Holm-adjusted significance levels.

4.2.1.2 Internal reliability and construct checks - Verifiers

Cronbach's α for FI and $\mathrm{PI}_{\mathrm{rev}}$ (with α-if-deleted)
For verifiers, the FI and $\mathrm{PI}_{\mathrm{rev}}$ scales exhibited notably different levels of internal consistency. Cronbach's α was 0.762 for the FI scale, indicating acceptable reliability. However, the $\mathrm{PI}_{\mathrm{rev}}$ scale showed substantially lower consistency with $\alpha = 0.445$, suggesting questionable internal coherence. The α if item deleted diagnostics revealed considerable variation in item contributions: for the FI scale, values remained within a range of 0.72 to 0.79, while for $\mathrm{PI}_{\mathrm{rev}}$, they ranged more widely from 0.32 to 0.51. Notably, removing *Interoperability* (NFR11) would increase α to 0.792 for FI (a gain of 0.030) and to 0.508 for $\mathrm{PI}_{\mathrm{rev}}$ (a gain of 0.063), indicating this item's relatively weak alignment with the other items in both scales.

Table 4.3: Block-level reliability estimates (Cronbach's α) and α-if-deleted coefficients for the 13-item SQRI scales for verifiers.

Block	Item / Statistic	α-if-deleted	Change (Δ)
FI	**Overall Block Reliability (13 items)**	**0.7621**	—
	Highest consistency increase if deleted:		
	NFR11 – Interoperability	0.7924	+0.0303
	NFR9 – Decentralization	0.7821	+0.0200
	NFR5 – Compatibility	0.7594	−0.0027
	Highest consistency decrease if deleted:		
	NFR18 – Security	0.7278	−0.0343
	NFR15 – Protection	0.7262	−0.0359
	NFR21 – Transparency	0.7194	−0.0427
PI_{rev}	**Overall Block Reliability (13 items)**	**0.4450**	—
	Highest consistency increase if deleted:		
	NFR11 – Interoperability	0.5079	+0.0628
	NFR1 – Accessibility	0.4707	+0.0257
	NFR21 – Transparency	0.4655	+0.0205
	Highest consistency decrease if deleted:		
	NFR15 – Protection	0.3948	−0.0502
	NFR2 – Authenticity	0.3389	−0.1061
	NFR14 – Privacy	0.3171	−0.1279

Note. The table reports the overall reliability for both blocks (FI and PI_{rev}) and highlights items with the highest and lowest changes in α if deleted. Positive Δ values indicate that removing the item increases internal consistency; negative values indicate a decrease.

The α if item deleted diagnostics reveal that the removal of *Interoperability* (NFR11) would yield the most substantial reliability gain for both blocks, suggesting that this item contributes the least to internal coherence within the scale relative to the other items. For the FI block, *Decentralization* (NFR9) also shows a positive change (0.020), though more minor in magnitude. For the PI_{rev} block, *Accessibility* (NFR1) and *Transparency* (NFR21) exhibit moderate positive changes (0.026 and 0.021, respectively), indicating that their removal would modestly improve internal consistency.

The majority of items exhibit a reduction in reliability if deleted, with the most significant decreases in the FI block for *Transparency* (NFR21, −0.043), *Protection* (NFR15, −0.036), and *Security* (NFR18, −0.034), and in the PI_{rev} block for *Privacy* (NFR14, −0.128), *Authenticity* (NFR2, −0.106), and *Protection* (NFR15, −0.050), highlighting their integral roles in their respective constructs.

The notably low PI_{rev} reliability ($\alpha = 0.445$) warrants particular attention, suggesting that verifiers may interpret problem scenarios less consistently than abstract FI. The substantial negative changes associated with *Privacy* (NFR14) and *Authenticity* (NFR2)

in PI_{rev} indicate these items are particularly essential to whatever coherence exists in this scale.

Inter-item structure: Spearman correlations

With 13 items per block, this analysis yielded $\binom{13}{2} = 78$ unique off-diagonal pairs per correlation matrix. The correlation coefficients and Holm-adjusted significance levels are reported in Excel sheets [3] and visualized in Figures 4.3 and 4.4.

FI block correlations

For the FI block, the average off diagonal correlation coefficient was $\bar{\rho} = 0.233$. Four of the 78 item pairs (5.1%) remained statistically significant after adjustment at the .05 level. The strongest observed association emerged between NFR21 (*Transparency*) and NFR18 (*Security*) at $\rho = 0.778$, $p_{adj} < .001$, followed by NFR14 (*Privacy*) and NFR20 (*Standard*) at $\rho = 0.699$, $p_{adj} < .01$. Two additional significant correlations were observed: NFR6 (*Consent*) with NFR14 (*Privacy*) at $\rho = 0.645$, $p_{adj} < .05$, and NFR15 (*Protection*) with NFR8 (*Cost*) at $\rho = 0.626$, $p_{adj} < .05$. The most negative association was between NFR21 (*Transparency*) and NFR9 (*Decentralization*) at $\rho = -0.203$, though this did not survive adjustment.

[3] For access to the Excel files, see Appendix B.

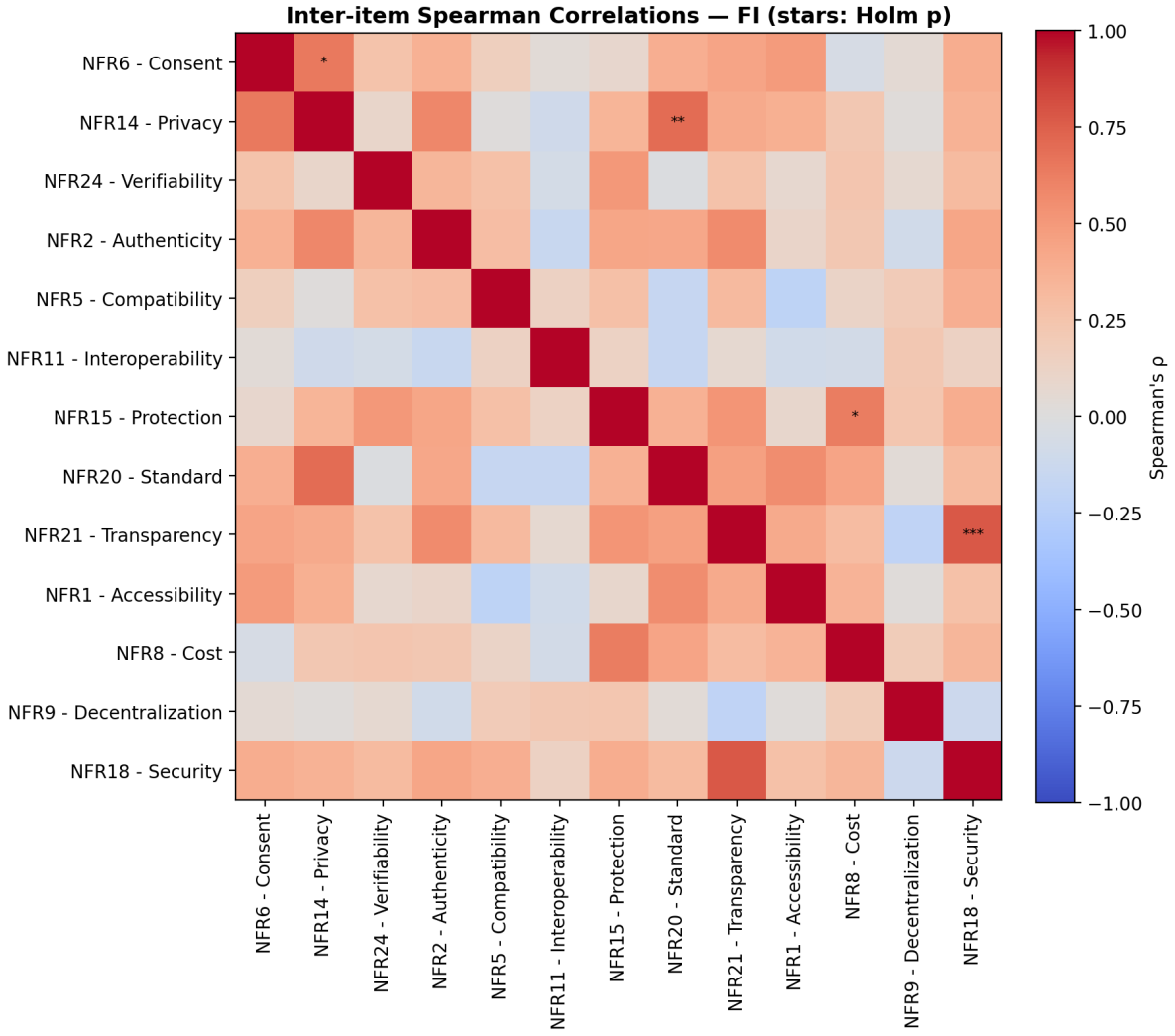

Figure 4.3: Spearman correlation heatmap for the FI block for verifiers, showing ρ values with adjusted significance levels.

$\mathbf{PI_{rev}}$ block correlations

The PI_{rev} block displayed lower average correlation strength with $\bar{\rho} = 0.093$. None of the 78 item pairs remained statistically significant after adjustment at the .05 level ($p_{\text{adj}} > .05$ for all pairs). The strongest association was observed between NFR14 (*Privacy*) and NFR2 (*Authenticity*) at $\rho = 0.533$, followed by NFR2 (*Authenticity*) and NFR15 (*Protection*) at $\rho = 0.466$. Notable positive correlations also included NFR11 (*Interoperability*) with NFR9 (*Decentralization*) at $\rho = 0.436$. The most negative correlation was between NFR21 (*Transparency*) and NFR9 (*Decentralization*) at $\rho = -0.352$, followed by NFR11 (*Interoperability*) with NFR20 (*Standard*) at $\rho = -0.308$. The absence of significant correlations after adjustment suggests that verifiers experience problems more heterogeneously than they assess abstract FI.

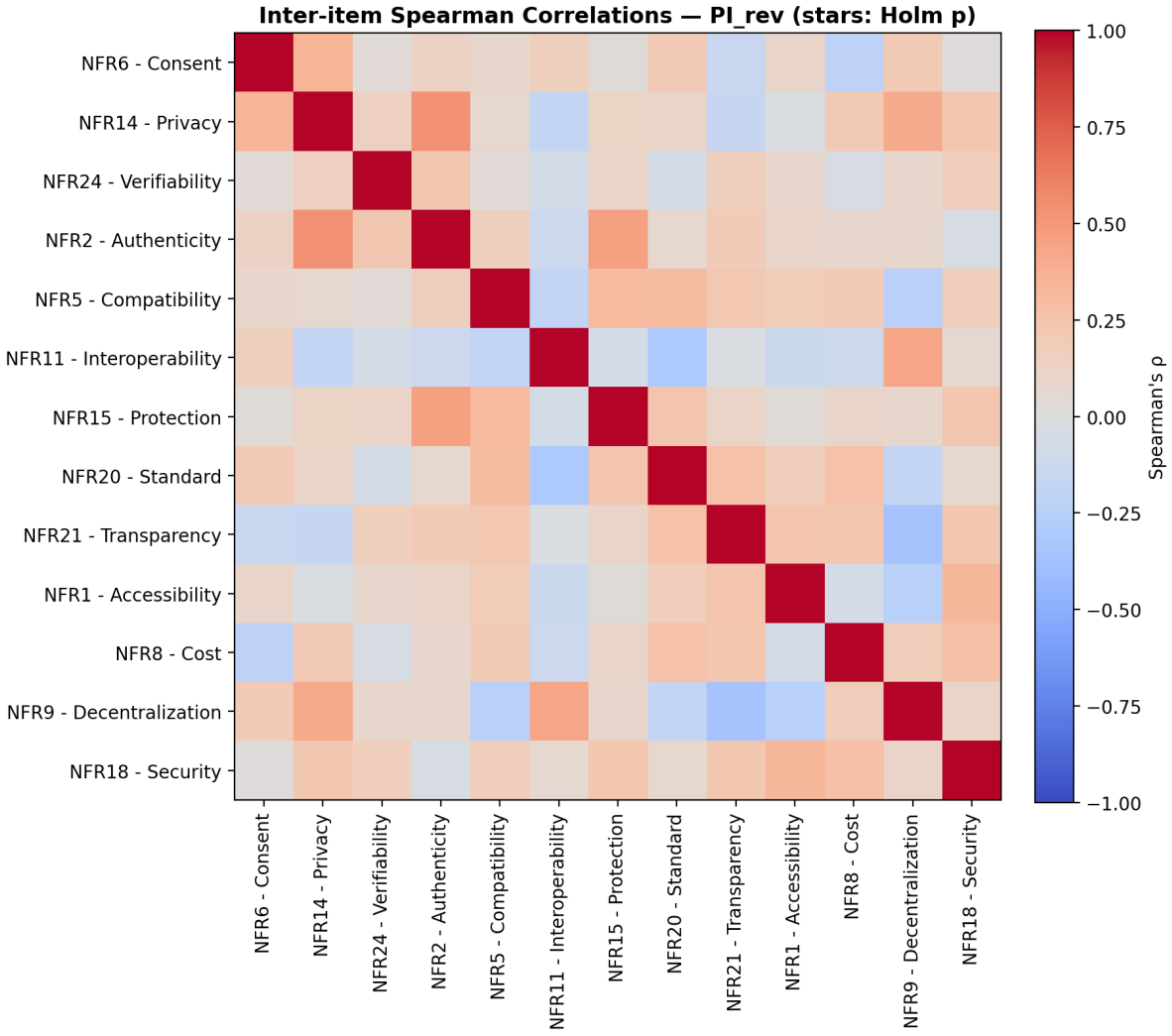

Figure 4.4: Spearman correlation heatmap for the PI_{rev} block for verifiers, showing ρ values with adjusted significance levels.

4.2.1.3 Internal reliability and construct checks - Issuers

Cronbach's α for FI and PI_{rev} (with α-if-deleted)
For issuers, both the FI and PI_{rev} scales exhibited moderate to low internal consistency. Cronbach's α was 0.584 for the FI scale and 0.472 for the PI_{rev} scale, indicating reliability substantially below conventional benchmarks. Furthermore, removing individual items revealed varied effects on overall consistency. For the FI scale, α if item deleted ranged from 0.526 to 0.604, while for the PI_{rev} scale, the range was broader, from 0.384 to 0.573.

In the FI block, the item corresponding to *Authenticity* (NFR2) was the only one whose removal noticeably raised α (to 0.604), suggesting it was somewhat less in sync with the

others; however, even this improvement remained modest. In the PI_{rev} block, removing *Standardization* (NFR20) yielded the most substantial reliability gain (to 0.573, $\Delta = 0.100$), indicating this item contributed the least to internal coherence. The majority of items, when removed, decreased reliability, with the most significant decreases associated with *Cost* (NFR8), *Verifiability* (NFR24), and *Security* (NFR18) in the FI block, and *Decentralization* (NFR9) and *Privacy* (NFR14) in the PI_{rev} block.

Table 4.4: Block-level reliability estimates (Cronbach's α) and α-if-deleted coefficients for the 12-item SQRI scales for issuers.

Block	Item / Statistic	α-if-deleted	Change (Δ)
FI	**Overall Block Reliability (12 items)**	**0.5838**	—
	Highest consistency increase if deleted:		
	NFR2 – Authenticity	0.6044	+0.0206
	Highest consistency decrease if deleted:		
	NFR8 – Cost	0.5256	−0.0583
	NFR24 – Verifiability	0.5409	−0.0429
	NFR18 – Security	0.5430	−0.0408
PI_{rev}	**Overall Block Reliability (12 items)**	**0.4721**	—
	Highest consistency increase if deleted:		
	NFR20 – Standard	0.5726	+0.1005
	NFR6 – Consent	0.4974	+0.0253
	Highest consistency decrease if deleted:		
	NFR9 – Decentralization	0.3843	−0.0878
	NFR14 – Privacy	0.3960	−0.0761
	NFR8 – Cost	0.4345	−0.0376

Note. The table reports the overall reliability for both blocks (FI and PI_{rev}) and highlights items with the highest and lowest changes in α if deleted. Positive Δ values indicate that removing the item increases internal consistency; negative values indicate a decrease.

The α if item deleted diagnostics reveal that the removal of *Authenticity* (NFR2) in the FI block would yield the most substantial reliability gain ($\Delta = 0.021$), though the improvement is minimal. In the PI_{rev} block, *Standardization* (NFR20) shows the largest potential improvement ($\Delta = 0.100$), suggesting it is less coherent with other items in the PI context.

Only minor positive changes are observed for a few additional items. In the FI block, *Privacy* (NFR14) shows virtually no change when removed ($\Delta \approx 0.000$), while in the PI_{rev} block, *Consent* (NFR6) shows a modest increase ($\Delta = 0.025$). All other items, when deleted, reduce internal consistency.

The majority of items exhibit a reduction in reliability if deleted, with the most significant decreases associated with core attributes such as *Cost* (NFR8, $\Delta = -0.058$), *Verifiability* (NFR24, $\Delta = -0.043$), and *Security* (NFR18, $\Delta = -0.041$) in the FI block, alongside

Decentralization (NFR9, $\Delta = -0.088$), *Privacy* (NFR14, $\Delta = -0.076$), and *Cost* (NFR8, $\Delta = -0.038$) in the PI_{rev} block. These findings highlight the integral nature of these items to their respective constructs, despite the overall low reliability of the measures.

Inter-item structure: Spearman correlations

Pairwise Spearman rank correlations (ρ) were computed for all SQRI items within each block (FI and PI_{rev}) to examine the internal structure and interrelationships among the 12 NFRs for issuers. With 12 items per block, this analysis yielded $\binom{12}{2} = 66$ unique off diagonal pairs per correlation matrix. Family wise Type I error was controlled within each block using the step down adjustment procedure, ensuring that the adjusted significance threshold accounts for multiple comparisons across all item pairs. The correlation coefficients, raw p values, adjusted p values, and significance markers are reported in the complete matrices provided in Excel sheets [4]. Figures 4.5 and 4.6 present heatmaps overlaid with significance stars, where ******* denotes $p_{\text{adj}} < .001$, ****** denotes $p_{\text{adj}} < .01$, and ***** denotes $p_{\text{adj}} < .05$, providing a visual summary of both correlation magnitude and statistical reliability across all pairs.

FI block correlations

For the FI block, correlations among the 12 quality requirement items displayed a mixed pattern with both positive and negative associations. The average off diagonal correlation coefficient was $\bar{\rho} = 0.123$, and only 1 of the 66 item pairs (1.5%) remained statistically significant after adjustment at the .05 level. The strongest observed association emerged between NFR15 (*Protection*) and NFR18 (*Security*), with $\rho = 0.701$, $p_{\text{adj}} < .05$, reflecting the conceptual overlap between these two core security oriented requirements in credential issuance contexts. The following highest correlations were observed between NFR21 (*Transparency*) and NFR24 (*Verifiability*) at $\rho = 0.503$, and between NFR21 (*Transparency*) and NFR8 (*Cost*) at $\rho = 0.465$. However, neither survived adjustment. These associations suggest that issuers who prioritize *Transparency* may also value *Verifiability* and *Cost* considerations, though the relationships are not uniformly strong across the sample.

The most negative association was between NFR14 (*Privacy*) and NFR8 (*Cost*), with a correlation coefficient of $\rho = -0.202$. However, this correlation did not survive adjustment ($p_{\text{adj}} > 0.05$). Other notable negative correlations included NFR6 (*Consent*) with NFR2 (*Authenticity*) at $\rho = -0.173$, and NFR11 (*Interoperability*) with NFR9 (*Decentralization*) at $\rho = -0.149$. These divergent ratings suggest potential trade-offs or organizational differences in how issuers balance *Privacy* versus resource constraints, consent management versus authentication priorities, and *Interoperability* versus decentralized architecture preferences.

Using conventional benchmarks for rank correlations (absolute ρ of .10 equals small, .30 equals moderate, .50 or more equals large), the observed effects in the FI block span from trivial to large. The mean off diagonal correlation $\bar{\rho} = .123$ indicates small typical associations. The strongest pair, *Protection* and *Security* (NFR15 and NFR18), exhibits

[4] see Appendix B.

a significant association with a correlation coefficient of $\rho = 0.70$. The weak negative pair *Privacy* and *Cost* (NFR14 and NFR8) at $\rho = -0.20$ has a small magnitude. Among the 66 pairs, the vast majority show weak associations, with only one reaching statistical significance after correction for multiple comparisons.

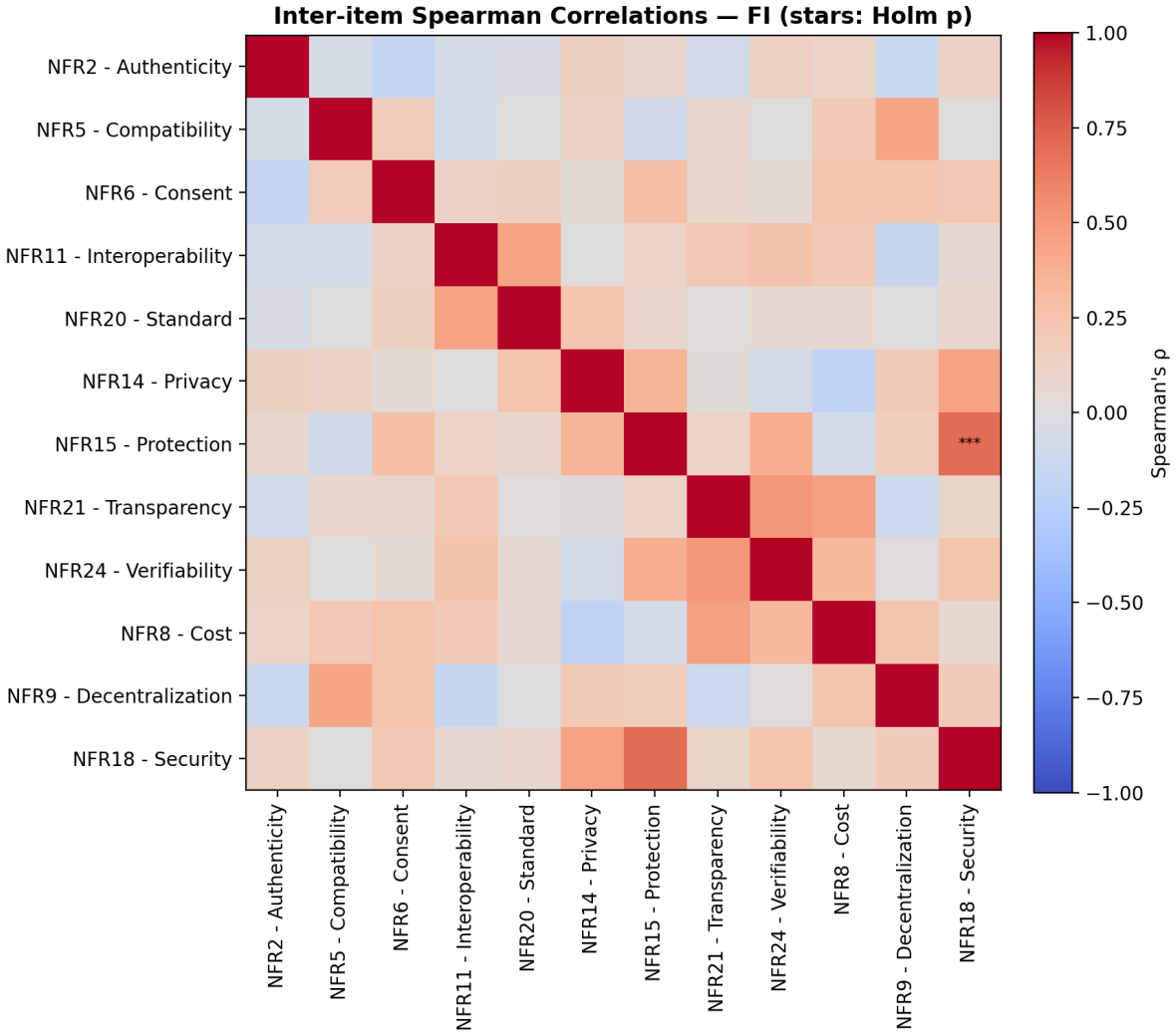

Figure 4.5: Spearman correlation heatmap for the FI block for issuers, showing ρ values with adjusted significance levels.

$\mathbf{PI_{rev}}$ block correlations

The $\mathrm{PI_{rev}}$ block displayed a similar overall pattern but with slightly lower average correlation strength. The mean off diagonal coefficient was $\bar{\rho} = 0.093$, and 2 of the 66 pairs (3.0%) were statistically reliable after adjustment. The strongest association was observed between NFR2 (*Authenticity*) and NFR24 (*Verifiability*), yielding $\rho = 0.555$, $p_{\mathrm{adj}} < .05$, reflecting issuers' perception that *Authenticity* and *Verifiability* are closely linked problem

areas in credential systems. The second significant correlation was between NFR15 (*Protection*) and NFR24 (*Verifiability*) at $\rho = 0.552$, $p_{\text{adj}} < .05$, further reinforcing the tight coupling between security related attributes in the problem space. Other notable positive correlations included NFR2 (*Authenticity*) with NFR15 (*Protection*) at $\rho = 0.460$ and NFR15 (*Protection*) with NFR18 (*Security*) at $\rho = 0.475$, though these did not survive adjustment.

A subset of negative associations was present in the PI_{rev} block, more pronounced than in FI. The most negative correlation was between NFR5 (*Compatibility*) and NFR24 (*Verifiability*), with $\rho = -0.371$, which did not reach adjusted significance. Other notable negative correlations included NFR20 (*Standard*) with NFR24 (*Verifiability*) at $\rho = -0.340$ and NFR20 (*Standard*) with NFR14 (*Privacy*) at $\rho = -0.334$. These negative associations suggest potential divergence in how issuers experience current problems, with some organizations facing compatibility issues. In contrast, others struggle with verifiability or experience trade-offs between standards and privacy requirements.

Applying the same benchmarks to the PI_{rev} block, the mean off diagonal correlation $\bar{\rho} = .093$ is small on average. The strongest pair, *Authenticity* and *Verifiability* (NFR2 and NFR24), has a correlation coefficient of $\rho = .555$, indicating a significant effect. The negative pair *Compatibility* and *Verifiability* (NFR5 and NFR24) at $\rho = -0.371$ is moderate in absolute magnitude. Of the 2 significant pairs after adjustment, both involve *Verifiability* (NFR24), highlighting its role as a central concern that correlates with multiple other problem areas.

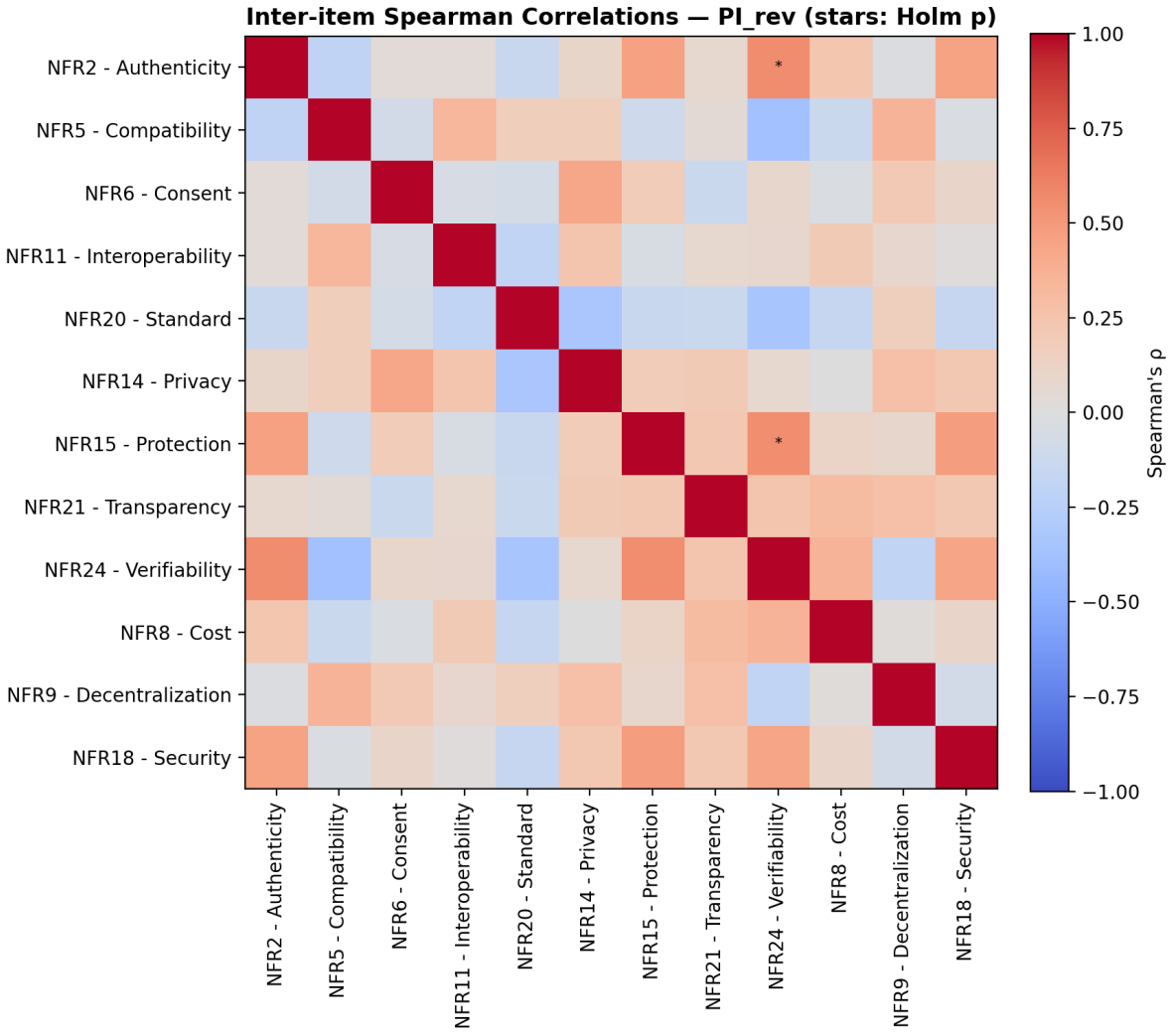

Figure 4.6: Spearman correlation heatmap for the $\mathrm{PI}_{\mathrm{rev}}$ block for issuers, showing ρ values with adjusted significance levels.

4.2.2 SQRI Analysis

The SQRI pipeline processed the Likert responses for FI and PI. Negatively worded PI items were reverse coded to PI_{rev} so that higher values consistently indicated greater importance or concern.

For each item (FI and PI_{rev}), the script computed means, medians, standard deviations, and interquartile ranges. All tables and publication ready figures were exported to an Excel workbook and graphics bundle (B).

For each item, the median was tested against the neutral point (3 on a 1 to 5 scale) using the Wilcoxon signed rank test. The analysis reported W and, where available, z, adjusted p values, and effect size r. A Sign test was included as a robustness check. Results were listed item wise in the exports.

To combine importance and problem concern, the script created a matrix with FI mean scores on the x axis and PI_{rev} mean scores on the y axis. The matrix divided items into four groups: High High priorities, High Low valued items, Low High latent risks, and Low Low items.

Between group differences (for example, gender, or occupation) were checked for each item using the Kruskal Wallis test. If a significant result was found, Dunn-Bonferroni post-hoc tests with adjustment for multiple comparisons were applied. Results included effect sizes, test statistics, adjusted p values, and summaries for each group.

To examine internal structure, the script computed Spearman rank correlations (for FI and PI_{rev}) and exported correlation matrices and heatmaps. Reliability (Cronbach's α) was reported in the dedicated reliability section (see 4.2.1); the corresponding values were also included in the exports.

All results were exported for reporting, including an Excel workbook (B) with descriptive statistics and test outcomes, as well as publication quality figures, such as the prioritization matrix and a correlation heatmap. Together with the earlier reliability evidence from Cronbach's alpha, these analyses provided a solid foundation for the Results chapter by identifying which quality requirements were rated as important and by showing that participants interpreted the items consistently.

4.2.2.1 Item-Level Descriptives & Rankings – Identity Holders (Users)

In the following section, descriptive statistics for each NFR were calculated on both the FI and PI_{rev} scales, including the mean, median, and standard deviation (SD) of the ratings. These statistics summarized the central tendency and variability of perceived importance for each quality attribute. To facilitate interpretation, each item was also ranked within the set of 24 NFRs based on its mean score on each scale. A rank of 1 indicated the highest mean importance (most important), and the highest rank number corresponded to the lowest importance relative to other items. Table 4.5 and Figures 4.7 and 4.8 presented the item level descriptive results and rankings side by side for the FI and PI_{rev} measures.

Table 4.5: Item–level descriptive statistics and importance rankings for the 24 NFR items (SQRI).

	FI				PI_{rev}			
NFR Item	**Mean**	**Median**	**SD**	**Rank**	**Mean**	**Median**	**SD**	**Rank**
NFR15 – Protection	4.62	5.0	0.84	1	4.70	5.0	0.75	1
NFR18 – Security	4.56	5.0	0.78	2	3.99	4.0	1.18	8
NFR16 – Recoverability	4.56	5.0	0.68	2	3.90	4.0	1.26	9
NFR7 – Control	4.51	5.0	0.85	3	4.23	5.0	1.06	4
NFR21 – Transparency	4.47	5.0	0.98	4	4.08	4.0	1.12	6
NFR1 – Accessibility	4.41	5.0	0.86	5	4.00	4.0	1.21	7
NFR22 – Usability	4.36	4.0	0.88	6	3.23	3.0	1.20	17
NFR5 – Compatibility	4.34	4.0	0.94	7	3.78	4.0	1.01	14
NFR2 – Authenticity	4.33	4.0	0.93	8	4.26	5.0	0.90	3
NFR14 – Privacy	4.29	5.0	0.92	9	3.86	4.0	1.04	10
NFR19 – Single Source	4.24	4.5	0.92	10	4.08	4.0	1.15	6
NFR12 – Persistence	4.24	4.0	0.88	10	3.80	4.0	1.29	12
NFR6 – Consent	4.23	5.0	1.01	11	4.38	5.0	1.13	2
NFR13 – Portability	4.20	4.0	0.99	12	3.78	4.0	1.07	14
NFR8 – Cost	4.17	4.0	0.94	13	4.08	4.0	1.08	6
NFR23 – User Experience	4.16	4.0	0.85	14	3.79	4.0	1.07	13
NFR4 – Availability	4.12	4.0	1.05	15	3.47	4.0	1.08	15
NFR20 – Standard	3.86	4.0	1.09	16	4.16	4.0	1.03	5
NFR3 – Autonomy	3.77	4.0	1.06	17	3.90	4.0	1.15	9
NFR9 – Decentralization	3.73	4.0	1.01	18	2.83	3.0	1.10	19
NFR11 – Interoperability	3.65	4.0	1.11	19	3.83	4.0	1.09	11
NFR10 – Existence	3.62	4.0	1.13	20	3.24	3.0	1.25	16
NFR17 – Representation	3.03	3.0	1.18	21	2.57	3.0	1.11	20
NFR24 – Verifiability	3.02	3.0	1.12	22	3.15	3.0	1.14	18

Note. For each NFR, the table reports the mean, median, and SD of the FI ratings and PI_{rev} ratings, along with the item's rank on each scale (where 1 indicates highest importance).

From these results, several clear patterns emerged regarding which NFRs respondents found most and least important. *Protection* (NFR15) stood out as the highest rated quality in both FI and PI_{rev}, with an average FI rating of $M = 4.62$ and an even higher PI_{rev} $M = 4.70$. This item had a median of 5 on both scales, indicating that at least half of the participants assigned it the maximum importance rating, which suggested a strong consensus on the critical importance of protection (that is, safeguarding identity data) in both general terms and when considering specific threat scenarios. Other top ranked NFRs on the FI scale included *Security* (NFR18) and *Recoverability* (NFR16), both with high FI scores ($M = 4.56$) and median 5, reflecting broad agreement on their importance. *Control* (NFR7) and *Transparency* (NFR21) were also among the top five FI items ($M = 4.51$ and $M = 4.47$, respectively), emphasizing the value users placed on having control over their identity data and on the system being transparent.

On the PI_{rev} scale, a broadly similar set of qualities was viewed as most critical, though with some differences in ordering. *Protection* (NFR15) remained the most vital issue (PI_{rev}

$M = 4.70$, rank 1), underscoring that a breach of data protection was the most alarming scenario for users. Following closely, *Consent* (NFR6) emerged as the second highest on the PI_{rev} scale ($M = 4.38$, median 5), despite being only mid ranked on FI, which suggested that while users might not always prioritize consent in the abstract, a scenario involving a consent violation (for example, the misuse of their personal data without permission) was highly concerning. *Authenticity* (NFR2) was another item with a high PI_{rev} rating ($M = 4.26$, rank 3), higher than its FI rank, suggesting that issues of identity authenticity (such as falsified or untrustworthy identity data) particularly resonated as serious problems. Meanwhile, qualities like *Control* (NFR7) and *Transparency* (NFR21) remained among the top tier in PI_{rev} as well (both with PI_{rev} $M > 4.0$), indicating consistency in their perceived importance across contexts.

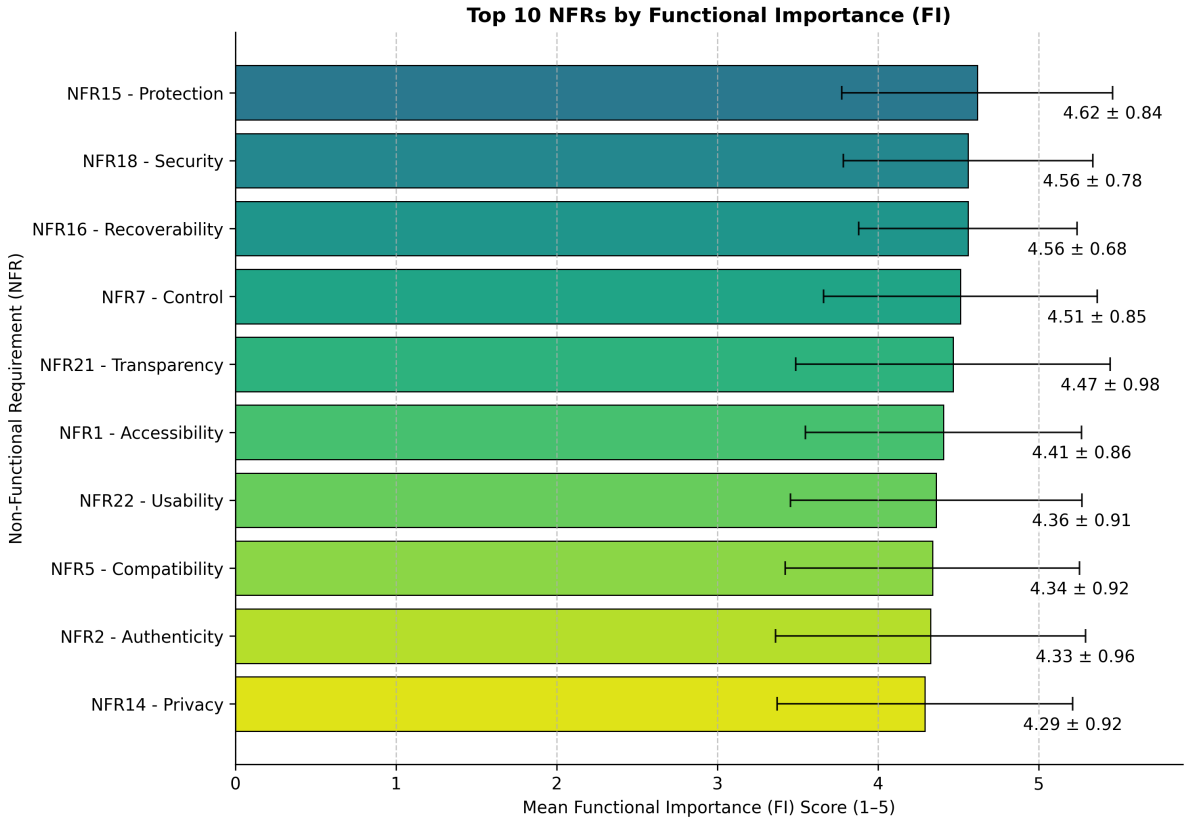

Figure 4.7: Top 10 NFRs by FI for users (mean ± SD, 1–5 scale).

At the lower end of the rankings, some NFRs consistently received comparatively lower importance scores. *Representation* (NFR17), which relates to how a user's identity is represented or formalized, was among the lowest on both scales, with an FI $M = 3.03$ (median 3) and an even lower PI_{rev} $M = 2.57$, which suggested that respondents were relatively neutral or divided on the importance of representation as a quality. *Decentralization* (NFR9) also scored low, particularly on the PI_{rev} measure (PI_{rev} $M = 2.83$, rank 19), indicating that the scenario of a system being overly centralized was not viewed as particularly critical by most users. *Verifiability* (NFR24) had the lowest FI $M = 3.02$ (rank 22) and a similarly low PI_{rev} $M = 3.15$, indicating that the ability to verify identity data, while somewhat important, was not a top priority relative to other qualities.

Notably, these lower ranked items typically had median ratings around 3 (the neutral midpoint of the scale), in contrast to the top items, which had medians of 5. This difference in medians highlighted that a majority of participants were ambivalent or split on the lesser items. In contrast, there was strong agreement on the importance of the top rated qualities.

To provide a comprehensive overview of all 24 NFRs, several key items merited explicit attention. *Accessibility* (NFR1) consistently held significant value, ranking prominently on both scales (FI $M = 4.41$, rank 5; PI_{rev} $M = 4.00$, rank 7), reflecting its alignment with the generally high regard for essential usability and control characteristics. Notably, some attributes displayed discrepancies between their perceived importance in the abstract and their relevance when framed as specific problems. For instance, *Autonomy* (NFR3) ranked in the lower to mid range for FI ($M = 3.77$, rank 17) but saw an increase in its importance under problem framing (PI_{rev} $M = 3.90$, rank 9), suggesting that its significance was heightened when viewed as a potential loss. Similarly, *Standard* (NFR20) showed a more pronounced shift from a mid ranked FI ($M = 3.86$, rank 16) to a higher position on PI_{rev} ($M = 4.16$, rank 5), indicating that the adverse effects of inadequate standardization were more noticeable, even if it was not a frequent concern in abstract discussions. A somewhat milder trend could be observed with *Interoperability* (NFR11), which had a relatively low FI ($M = 3.65$, rank 19) but improved under problem framing (PI_{rev} $M = 3.83$, rank 11).

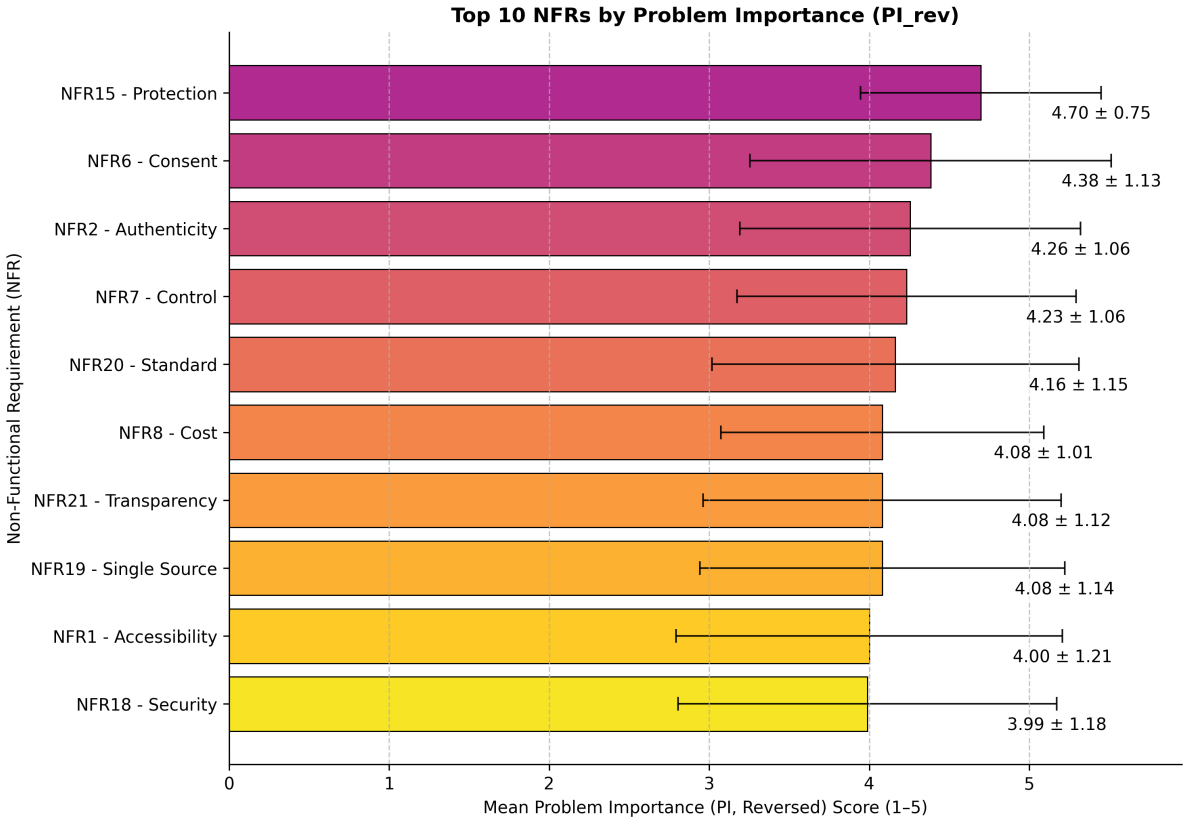

Figure 4.8: Top 10 NFRs by PI_{rev} for users (mean ± SD, 1–5 scale).

Certain attributes remained fundamentally important yet received less attention when framed as specific issues. *Compatibility* (NFR5) and *Portability* (NFR13) both occupied

the upper mid range for Favorability Index (FI) scores ($M = 4.34$, rank 7; and $M = 4.20$, rank 12, respectively) but dropped to mid lower ranks on the PI_{rev} scale ($M = 3.78$, rank 14 for both). *Usability* (NFR22) exemplified this trend most strikingly, ranking high on the FI ($M = 4.36$, rank 6) while scoring much lower on the PI_{rev} ($M = 3.23$, rank 17). This suggested that, although usability was recognized as valuable, the specific problem scenario presented did not resonate as critical compared to risks such as protection or consent violations. In contrast, *Cost* (NFR8) became more salient when viewed as a problem (PI_{rev} $M = 4.08$, rank 6) while maintaining an upper mid score on the FI ($M = 4.17$, rank 13), indicating that cost related concerns were perceived as significant trade-offs.

A set of attributes clustered around the upper mid range on FI with middling PI_{rev} positions: *Persistence* (NFR12) (FI $M = 4.24$, rank 10; PI_{rev} $M = 3.80$, rank 12), *User Experience* (NFR23) (FI $M = 4.16$, rank 14; PI_{rev} $M = 3.79$, rank 13), and *Single Source* (NFR19), which remained relatively strong across both perspectives (FI $M = 4.24$, rank 10; PI_{rev} $M = 4.08$, rank 6). *Availability* (NFR4) occupied the middle on both scales (FI $M = 4.12$, rank 15; PI_{rev} $M = 3.47$, rank 15). *Privacy* (NFR14) was consistently upper tier (FI $M = 4.29$, rank 9; PI_{rev} $M = 3.86$, rank 10), reinforcing the prominence of protection and security adjacent concerns already observed. At the lower end, *Existence* (NFR10) remained subdued across both views (FI $M = 3.62$, rank 20; PI_{rev} $M = 3.24$, rank 16).

4.2.2.2 Friedman Rank Tests – Identity Holders (Users)

The set of NFR items was treated as a repeated measures factor and analyzed using the Friedman test to identify overall differences in priorities. Kendall's W was reported as an index of agreement. When the result was significant, Dunn-Bonferroni pairwise comparisons were performed to examine differences in mean ranks, with effect sizes calculated as $r = |z|/\sqrt{n}$. NFRs were then ranked by their Friedman mean ranks, and only the five most considerable pairwise differences were highlighted. Heatmaps with stars for significance levels and color coded effect sizes were used to visualize where contrasts between NFRs were most pronounced. This approach provided a non parametric, stakeholder specific view of NFR priorities, robust for Likert data and directly comparable between the FI and PI_{rev} blocks.

The Friedman test for the identity holders showed significant differences in how users prioritized NFRs. For FI, the test yielded $\chi^2(23) = 375.59$, $p = 1.84 \times 10^{-65}$, $W = 0.190$, $n = 86$. For PI_{rev}, results were $\chi^2(23) = 432.39$, $p = 3.71 \times 10^{-77}$, $W = 0.219$, $n = 86$. Both Kendall's W values indicated moderate and statistically robust agreement among respondents, with slightly stronger concordance for PI_{rev} than for FI.

Table 4.6 presented the Friedman mean ranks for all 24 NFRs in the FI and PI_{rev} blocks from the users sample. Each mean rank was based on assigning a rank from 1 to 24 for each NFR per respondent, then averaging these ranks across all participants. The scale ranged from 1 (lowest priority) to 24 (highest priority). *Protection* ranked highest in both blocks, with a mean rank of 16.35 in FI and 18.16 in PI_{rev}. *Security*, *Recoverability*, *Control*, and *Transparency* followed as top priorities in FI, while *Consent* and *Authenticity*

joined the upper tier in PI_{rev}. *Representation* and *Verifiability* had the lowest mean ranks in both blocks, indicating the lowest priority among users.

Table 4.6: Friedman mean ranks by NFR for Users (FI and PI_{rev} blocks).

FI Rank	Code	Name	FI MeanRank	PI_{rev} MeanRank	PI_{rev} Rank
1	NFR15	Protection	16.35	18.16	1
2	NFR18	Security	15.47	13.65	10
3	NFR16	Recoverability	15.38	12.37	12
4	NFR7	Control	15.22	15.06	4
5	NFR21	Transparency	15.00	14.04	8
6	NFR1	Accessibility	14.27	13.78	9
7	NFR22	Usability	14.05	8.90	20
8	NFR2	Authenticity	13.82	15.26	3
9	NFR5	Compatibility	13.62	12.34	14
10	NFR14	Privacy	13.44	9.09	18
11	NFR12	Persistence	13.03	12.37	12
12	NFR6	Consent	13.33	16.47	2
13	NFR19	Single Source	12.90	14.26	6
14	NFR4	Availability	12.34	9.97	15
15	NFR23	User Experience	12.13	9.19	17
16	NFR20	Standard	11.08	14.80	5
17	NFR3	Autonomy	10.03	12.52	11
18	NFR13	Portability	10.03	12.37	13
19	NFR8	Cost	9.97	14.08	7
20	NFR9	Decentralization	9.78	7.06	22
21	NFR10	Existence	9.77	9.29	16
22	NFR11	Interoperability	9.72	8.98	19
23	NFR24	Verifiability	6.87	8.28	21
24	NFR17	Representation	6.53	6.30	24

Note. FI and PI_{rev} mean ranks are reported for all 24 NFRs, ordered by FI rank. Higher mean rank means higher relative priority; PI_{rev} ranks are also provided for each item for comparison.

Table 4.7 shows the five strongest and weakest significant pairwise differences for each block in the users sample, with the direction indicating which NFR had the higher mean rank. Post-hoc pairwise comparisons revealed significant differences in 70 out of 276 NFR pairs (25.4%) for the FI block and in 97 out of 276 pairs (35.1%) for the PI_{rev} block. In the FI block, *Protection versus Representation* ($r = 0.981$) and *Protection versus Verifiability* ($r = 0.948$) are the strongest contrasts, while the weakest include *Authenticity versus Decentralization* and *Standard versus Verifiability* (r values between 0.404 and 0.421). In the PI_{rev} block, *Protection* ranks strongly above *Representation* ($r = 1.186$) and *Decentralization* ($r = 1.110$), with additional strong pairs involving *Consent*. The weakest significant pairs here are *Availability versus Transparency* and *Portability versus Verifiability* (r values near 0.41). Adjusted p-values are given for each pair to show statistical significance.

Table 4.7: Top five strongest and weakest Bonferroni-significant pairwise contrasts by block (Users).

Block	Type	NFR Pair	Pair (names)	r	p_{adj}
FI	Strongest	NFR15 – NFR17	Protection > Representation	0.981	$< 10^{-16}$
FI	Strongest	NFR15 – NFR24	Protection > Verifiability	0.948	$< 10^{-16}$
FI	Strongest	NFR17 – NFR18	Representation > Security	0.893	6.13×10^{-14}
FI	Strongest	NFR16 – NFR17	Recoverability > Representation	0.884	6.13×10^{-14}
FI	Strongest	NFR7 – NFR17	Control > Representation	0.869	2.45×10^{-13}
FI	Weakest	NFR2 – NFR9	Authenticity > Decentralization	0.404	0.0494
FI	Weakest	NFR10 – NFR2	Existence > Authenticity	0.405	0.0473
FI	Weakest	NFR2 – NFR11	Authenticity > Interoperability	0.410	0.0389
FI	Weakest	NFR7 – NFR20	Control > Standard	0.415	0.0334
FI	Weakest	NFR20 – NFR24	Standard > Verifiability	0.421	0.0262
PI_{rev}	Strongest	NFR15 – NFR17	Protection > Representation	1.186	$< 10^{-16}$
PI_{rev}	Strongest	NFR15 – NFR9	Protection > Decentralization	1.110	$< 10^{-16}$
PI_{rev}	Strongest	NFR6 – NFR17	Consent > Representation	1.017	$< 10^{-16}$
PI_{rev}	Strongest	NFR15 – NFR24	Protection > Verifiability	0.988	$< 10^{-16}$
PI_{rev}	Strongest	NFR6 – NFR9	Consent > Decentralization	0.941	$< 10^{-16}$
PI_{rev}	Weakest	NFR4 – NFR21	Availability > Transparency	0.408	0.0434
PI_{rev}	Weakest	NFR15 – NFR8	Protection > Cost	0.408	0.0424
PI_{rev}	Weakest	NFR13 – NFR24	Portability > Verifiability	0.409	0.0406
PI_{rev}	Weakest	NFR6 – NFR13	Consent > Portability	0.410	0.0398
PI_{rev}	Weakest	NFR4 – NFR8	Availability > Cost	0.412	0.0372

Note. Only the five strongest and five weakest significant pairs are reported for each block. Directionality reflects the higher mean rank for the first NFR; r values are effect sizes for difference, and p_{adj} are adjusted for family-wise error.

The heatmaps show all pairwise comparisons between NFR items for each stakeholder group, with cell color indicating effect size and stars indicating statistical significance based on adjusted p-values.

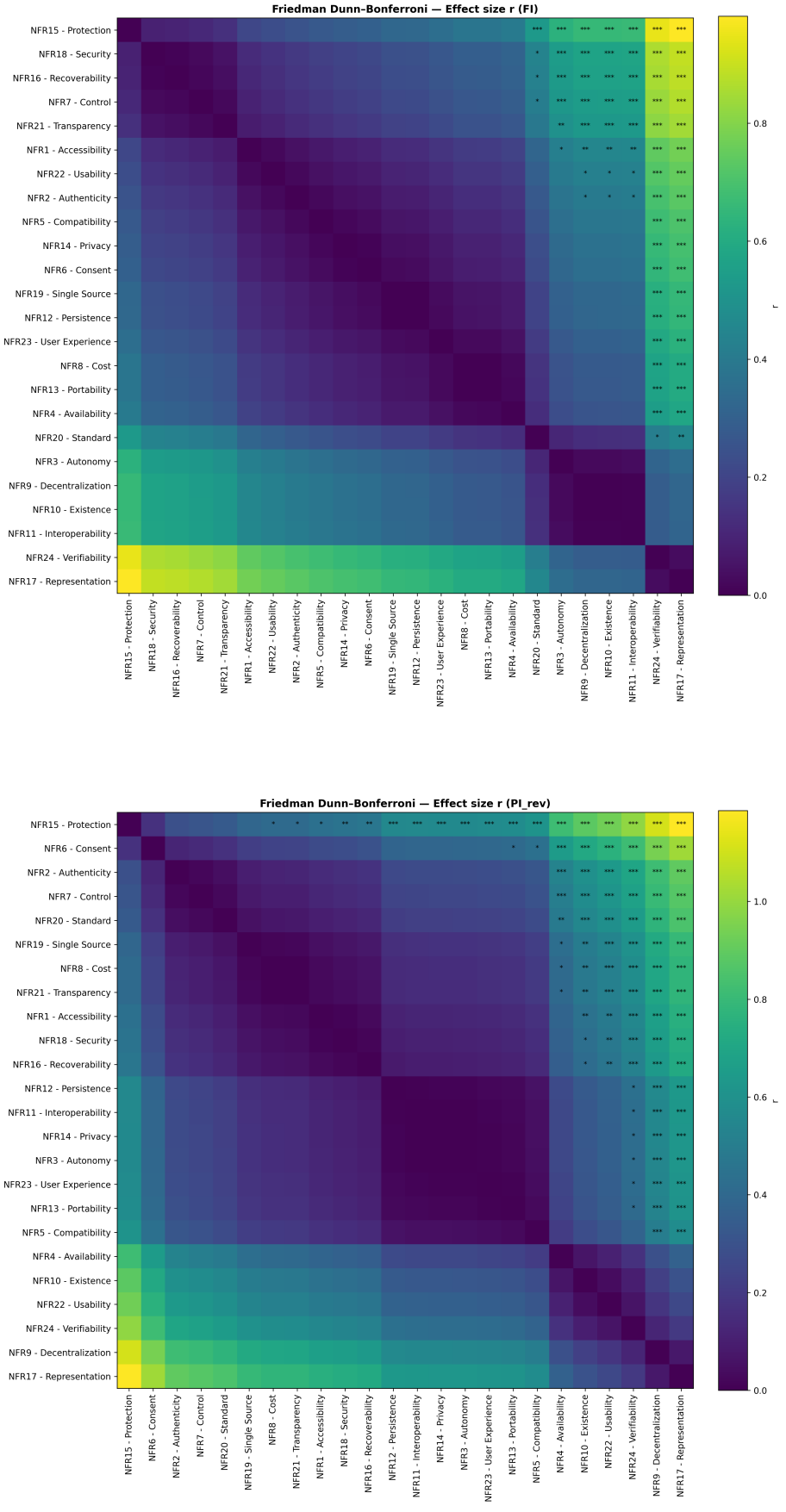

Figure 4.9: Friedman post-hoc effect-size heatmaps.

4.2.2.3 Prioritization matrix ($\mathbf{FI_{mean}} \times \mathbf{PI}_{rev,\mathbf{mean}}$) – Identity Holders (Users)

A two-dimensional scatterplot matrix combines FI and PI into a single prioritization view, with mean FI on the x-axis and mean PI_{rev} on the y-axis. Quadrant thresholds are set to the within-role medians (FI = 4.24, PI_{rev} = 3.88) to mitigate ceiling effects typical of 1–5 Likert data; dashed crosshairs at the means (FI = 4.10, PI_{rev} = 3.80) are shown only as references. The matrix is divided into four zones: High FI / High PI (top-right), High FI / Low PI (top-left), Low FI / High PI (bottom-right), and Low FI / Low PI (bottom-left). Figure 4.10 shows the result.

High FI / High PI: Top-priority attributes
Eight items are classified in the top-right quadrant, representing requirements that users consider both functionally important and comparatively well delivered. This quadrant covers FI scores from 4.24 to 4.62 and PI_{rev} scores from 3.90 to 4.70. The average scores within this quadrant are FI = 4.46 and PI_{rev} = 4.15.

Table 4.8: High FI / High PI quadrant: Top-priority attributes (n = 8)

Code	NFR Item	FI Mean	PI_{rev} Mean
NFR15	Protection	4.62	4.70
NFR7	Control	4.51	4.23
NFR2	Authenticity	4.33	4.26
NFR1	Accessibility	4.41	4.00
NFR19	Single Source	4.24	4.08
NFR21	Transparency	4.47	4.08
NFR18	Security	4.56	3.99
NFR16	Recoverability	4.56	3.90

High FI / Low PI quadrant: Important in principle; Lower problem salience
Four items score above the FI median but below the PI_{rev} median. The average scores for these items are FI = 4.34 and PI_{rev} = 3.67.

Table 4.9: High FI / Low PI quadrant: Important in principle; Lower problem salience (n = 4)

Code	NFR Item	FI Mean	PI_{rev} Mean
NFR22	Usability	4.36	3.23
NFR5	Compatibility	4.34	3.78
NFR14	Privacy	4.29	3.86
NFR12	Persistence	4.24	3.79

Low FI / High PI quadrant: Over-delivered attributes
Four items fall below the FI median but above the PI_{rev} median (FI = 3.95 on average; PI_{rev} = 4.10).

Table 4.10: Low FI / High PI quadrant: Over-delivered attributes (n = 4)

Code	**NFR Item**	**FI Mean**	**PI_{rev} Mean**
NFR20	Standard	3.86	4.16
NFR3	Autonomy	3.77	3.90
NFR6	Consent	4.23	4.38
NFR8	Cost	4.17	4.08

Low FI / Low PI quadrant: Lowest-priority attributes
Eight items are positioned below the medians on both axes, with average scores of FI = 3.59 and PI_{rev} = 3.20.

Table 4.11: Low FI / Low PI quadrant: Lowest-priority attributes (n = 8)

Code	**NFR Item**	**FI Mean**	**PI_{rev} Mean**
NFR13	Portability	4.20	3.78
NFR23	User Experience	4.16	3.79
NFR4	Availability	4.12	3.47
NFR9	Decentralization	3.73	2.83
NFR11	Interoperability	3.65	3.83
NFR10	Existence	3.62	3.24
NFR17	Representation	3.03	2.57
NFR24	Verifiability	3.02	3.15

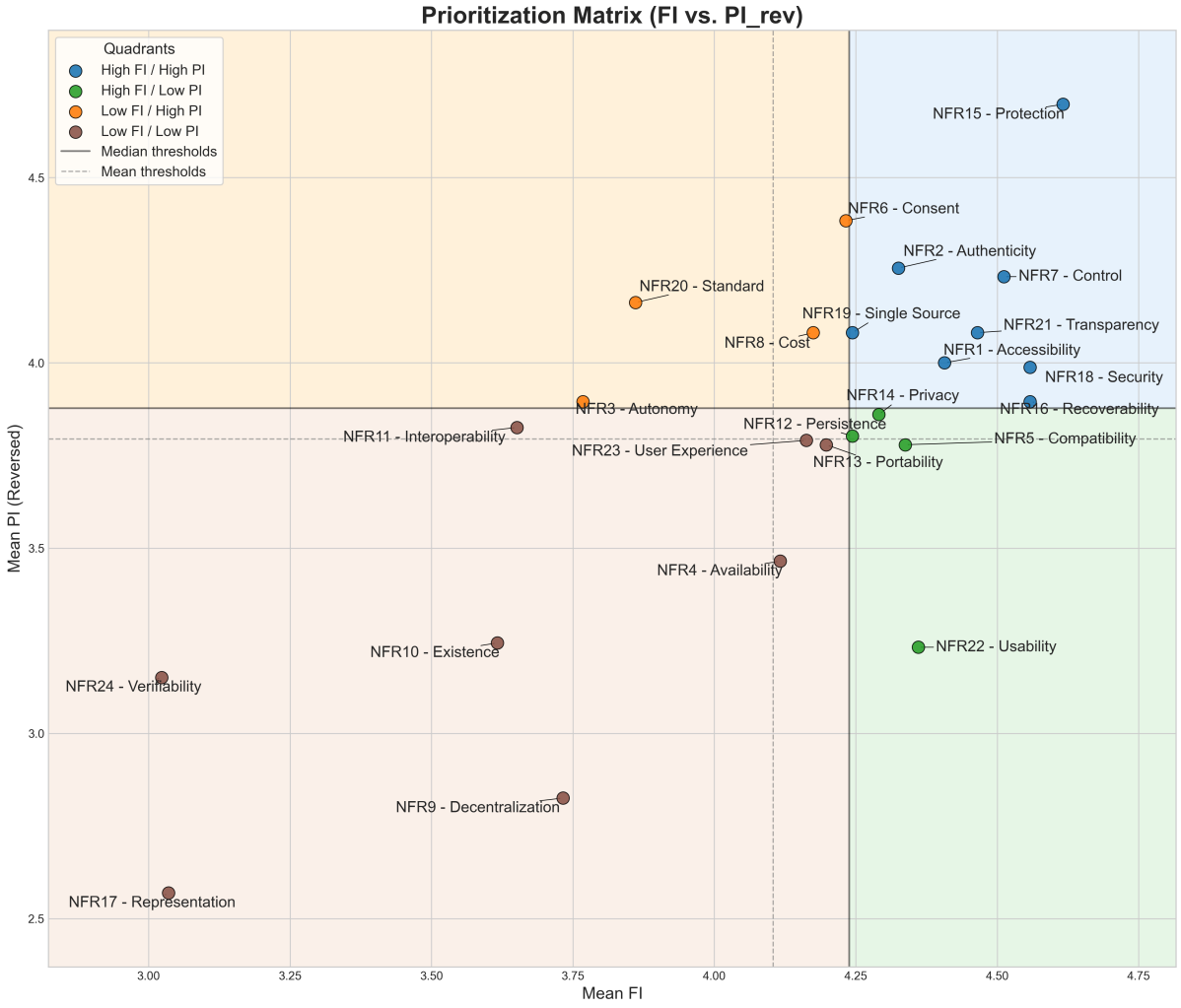

Figure 4.10: Prioritization matrix for the NFRs, displaying mean FI versus mean $\mathrm{PI}_{\mathrm{rev}}$. Shading and solid crosshairs indicate the median-based classification, while dashed crosshairs represent mean values for reference.

4.2.2.4 Group Differences Across Profession & Gender – Identity Holders (Users)

To assess whether NFR priorities varied systematically by respondent demographics, non-parametric Kruskal–Wallis rank-sum tests were conducted separately for each of the 24 items in both the FI and PI_{rev} blocks, comparing ratings across gender (three groups: Female, Male, No Answer) and professional role (four groups: In training, Manager/Executive, Other, Professional/Academic occupation).

The Kruskal–Wallis H statistic tests the null hypothesis that all groups are drawn from the same distribution, making it appropriate for ordinal Likert-scale data with unequal group sizes. For each comparison, the test statistic H, raw p-value, ε^2 effect size (ϵ^2, a rank-based measure of association strength ranging from 0 to 1), and Holm-adjusted p-value ($p_{\mathrm{adj.}}$) were computed to control the family-wise error rate within each comparison family. Complete test results, group-wise descriptive means, and post-hoc pairwise comparisons are available in the referenced repository dataset B.

Gender comparisons

Across the FI block, no items showed statistically significant gender differences after Holm correction ($p_{\text{adj.}} \geq .05$ for all 24 items, range 0.22–1.00). The five items with the smallest raw p-values (before correction) are displayed in Table 4.12.

Table 4.12: Kruskal-Wallis tests across gender (Female, Male, No Answer) for the FI block (Identity Holders, $n = 86$).

NFR item	H	p **(raw)**	ε^2	$p_{\textbf{adj.}}$	**Female**	**Male**	**No Answer**
Authenticity (NFR2)	9.40	.009	0.089	0.22	4.07	4.62	4.20
Portability (NFR13)	7.88	.019	0.071	0.45	3.93	4.46	4.40
Representation (NFR17)	7.06	.029	0.061	0.65	2.88	3.03	4.40
Verifiability (NFR24)	6.70	.035	0.057	0.74	2.67	3.31	3.80
User Experience (NFR23)	5.99	.050	0.048	1.00	3.90	4.41	4.40

Notes. The table lists the five items with the smallest *raw* p-values; none remained significant after Holm correction across 24 items. KW-tests across three gender groups; family-wise error controlled within the FI block using Holm adjustment over 24 items. After correction, no items were significant ($p_{\text{adj.}} \geq 0.05$; range 0.22–1.00). Across all 24 items, omnibus effect sizes were small on average (mean $\varepsilon^2 = 0.019$, range 0.0003–0.089).

Across the PI_{rev} block, no items showed significant gender differences after Holm correction ($p_{\text{adj.}} \geq .05$ for all 24 items, range 0.40–1.00). The five items with the smallest raw p-values are shown in Table 4.13.

Table 4.13: Kruskal–Wallis tests across gender (Female, Male, No Answer) for the PI_{rev} block (Identity Holders, $n = 86$).

NFR item	H	p **(raw)**	ε^2	$p_{\textbf{adj.}}$	**Female**	**Male**	**No Answer**
Interoperability (NFR11)	8.17	.017	0.074	0.40	3.81	4.03	2.40
Authenticity (NFR2)	6.70	.035	0.057	0.81	4.31	4.38	2.80
Autonomy (NFR3)	5.80	.055	0.046	1.00	3.83	4.13	2.60
Security (NFR18)	5.77	.056	0.045	1.00	4.07	4.08	2.60
Single Source (NFR19)	5.63	.060	0.044	1.00	4.05	4.28	2.80

Notes. The table lists the five items with the smallest *raw* p-values; none remained significant after Holm correction across 24 items. KW-tests across three gender groups; family-wise error controlled within the PI_{rev} block using Holm adjustment over 24 items. After correction, no items were significant ($p_{\text{adj.}} \geq .05$; range 0.40–1.00). Omnibus effects were small on average (mean $\varepsilon^2 = 0.022$, range 0.0006–0.074).

Profession comparisons

Across the FI block, no items differed significantly by professional role after Holm correction (all $p_{\text{adj.}} = 1.00$). The five items with the smallest raw p-values are shown in Table 4.14.

Table 4.14: Kruskal–Wallis tests across professional role (Training, Manager/Executive, Professional/Academic, Other) for the FI block (Identity Holders, $n = 86$).

NFR item	H	p **(raw)**	ε^2	$p_{\text{adj.}}$	**Training**	**Manager**	**Prof/Acad**	**Other**
Portability (NFR13)	8.08	.044	0.075	1.00	4.44	3.75	4.43	3.72
Protection (NFR15)	7.83	.050	0.071	1.00	4.56	4.75	5.00	4.11
Standard (NFR20)	7.61	.055	0.068	1.00	3.94	4.12	4.50	3.22
Authenticity (NFR2)	6.45	.092	0.051	1.00	4.56	3.62	4.57	4.17
Persistence (NFR12)	6.35	.096	0.049	1.00	4.28	4.50	4.57	3.78

Notes. Five items with the smallest *raw* p-values; none remained significant after Holm correction across 24 items (all $p_{\text{adj.}} = 1.00$). KW-tests across four role groups; Holm correction within the FI block across 24 items (all $p_{\text{adj.}} = 1.00$). Omnibus effects were small on average (mean $\varepsilon^2 = 0.020$, range 0.0001–0.075).

Across the PI_{rev} block, no items showed significant role differences after Holm correction ($p_{\text{adj.}} \geq .05$ for all 24 items, range 0.18–1.00). The five items with the smallest raw p-values are presented in Table 4.15.

Table 4.15: Kruskal–Wallis tests across professional role (Training, Manager/Executive, Professional/Academic, Other) for the PI_{rev} block (Identity Holders, $n = 86$).

NFR item	H	p **(raw)**	ε^2	$p_{\text{adj.}}$	**Training**	**Manager**	**Prof/Acad**	**Other**
Privacy (NFR14)	12.00	.007	0.132	0.18	3.66	4.12	4.57	3.50
Standard (NFR20)	11.00	.012	0.117	0.27	4.25	3.62	4.79	3.78
Representation (NFR17)	9.47	.024	0.095	0.52	2.88	2.88	2.57	1.94
Authenticity (NFR2)	8.68	.034	0.084	0.71	4.38	4.50	4.79	3.72
User Experience (NFR23)	6.94	.074	0.058	1.00	3.88	4.25	4.14	3.28

Notes. Five items with the smallest *raw* p-values; none remained significant after Holm correction across 24 items. KW-tests across four role groups; family-wise error controlled within the PI_{rev} block using Holm adjustment over 24 items ($p_{\text{adj.}}$ range 0.18–1.00). Mean omnibus effect size $\varepsilon^2 = 0.031$ (max 0.132 for NFR14).

SSI experience comparisons

Across the FI block, no items differed significantly by SSI experience after adjustment (all $p_{\text{adj.}} = 1.00$). The five items with the smallest raw p-values are shown in Table 4.16. Omnibus effects were very small on average ($\bar{\varepsilon}^2 = 0.003$, range 0.000 to 0.019, maximum for *Representation*, NFR17).

Table 4.16: Kruskal Wallis tests across SSI experience (Experienced vs. No Experience) for the FI block (Identity Holders, $n = 86$).

NFR item	H	p **(raw)**	ε^2	$p_{\textbf{adj.}}$	**No SSI exp.**	**SSI exp.**
Representation (NFR17)	2.59	.107	0.019	1.00	2.90	3.33
Availability (NFR4)	2.27	.132	0.015	1.00	3.98	4.41
Autonomy (NFR3)	1.97	.160	0.012	1.00	3.88	3.52
Verifiability (NFR24)	1.67	.197	0.008	1.00	3.15	2.74
Protection (NFR15)	1.49	.223	0.006	1.00	4.69	4.44

Notes. Five items with the smallest *raw* p-values; none remained significant after adjustment across 24 items (all $p_{\text{adj.}} = 1.00$). KW tests across two groups (no vs. some SSI experience); adjustment within the block across 24 items.

Across the PI_{rev} block, no items showed significant differences by SSI experience after adjustment (all $p_{\text{adj.}} = 1.00$). The five items with the smallest raw p-values are presented in Table 4.17. Omnibus effects were also very small on average ($\bar{\varepsilon}^2 = 0.002$, range 0.000 to 0.013, maximum for *User Experience*, NFR23).

Table 4.17: Kruskal Wallis tests across SSI experience (Experienced vs. No Experience) for the PI*rev* block (Identity Holders, $n = 86$).

NFR item	H	p **(raw)**	ε^2	$p_{\textbf{adj.}}$	**No SSI exp.**	**SSI exp.**
User Experience (NFR23)	2.12	.146	0.013	1.00	3.93	3.48
Security (NFR18)	1.95	.162	0.011	1.00	4.12	3.70
Consent (NFR6)	1.92	.166	0.011	1.00	4.51	4.11
Persistence (NFR12)	1.17	.279	0.002	1.00	3.71	4.00
Privacy (NFR14)	0.72	.397	0.000	1.00	3.93	3.70

Notes. Five items with the smallest *raw* p-values; none remained significant after adjustment across 24 items (all $p_{\text{adj.}} = 1.00$). KW tests across two groups (no vs. some SSI experience); adjustment within the block across 24 items.

Post-hoc pairwise comparisons

In accordance with standard hierarchical testing protocols, pairwise Dunn tests with Bonferroni-adjusted p-values and rank-biserial effect sizes were planned exclusively for items where the omnibus Kruskal–Wallis test remained significant after Holm adjustment at $\alpha = .05$. However, none of the 144 omnibus tests (24 items $\times$ 2 blocks $\times$ 3 grouping variables) met this criterion after family-wise error control; therefore, no post-hoc comparisons were conducted.

4.2.2.5 Item-Level Descriptives and Rankings – Verifiers

In the following section, descriptive statistics for each NFR were calculated on both the FI and PI_{rev} scales, including mean, median, and standard deviation. To facilitate interpretation, each item was ranked within the set of 13 NFRs based on its mean score on each scale. Table 4.18 and Figures 4.11 and 4.12 present the item-level descriptive results and rankings side by side for the FI and PI_{rev} measures.

Table 4.18: Item–level descriptive statistics and importance rankings for the 13 NFR items (SQRI) for verifiers.

	FI (Functional Importance)				**PI_{rev} (Problem Importance, reversed)**			
NFR Item	**Mean**	**Median**	**SD**	**Rank**	**Mean**	**Median**	**SD**	**Rank**
NFR14 – Privacy	4.41	5.0	1.05	1	3.78	4.0	1.42	7
NFR2 – Authenticity	4.26	4.0	0.94	2	3.85	4.0	1.32	5
NFR15 – Protection	4.26	5.0	1.10	3	4.30	5.0	0.95	1
NFR20 – Standard	4.26	5.0	1.02	4	3.89	4.0	1.05	4
NFR8 – Cost	4.15	5.0	1.10	5	3.81	4.0	1.00	6
NFR18 – Security	4.04	4.0	1.09	6	3.96	4.0	1.26	3
NFR24 – Verifiability	3.89	4.0	1.19	7	3.19	3.0	1.14	12
NFR6 – Consent	3.85	4.0	1.29	8	4.00	5.0	1.44	2
NFR1 – Accessibility	3.81	4.0	0.96	9	3.44	3.0	0.85	10
NFR21 – Transparency	3.74	4.0	1.20	10	3.63	4.0	1.21	8
NFR5 – Compatibility	3.52	4.0	1.25	11	3.52	4.0	1.12	9
NFR11 – Interoperability	3.33	4.0	1.33	12	3.15	3.0	1.32	13
NFR9 – Decentralization	3.11	3.0	1.19	13	3.22	3.0	1.12	11

Note. For each NFR, the table reports the mean, median, and standard deviation (SD) of the FI ratings and PI_{rev} ratings, along with the item's rank on each scale (where 1 indicates the highest importance).

From these results, several clear patterns emerged regarding which NFRs verifiers found most and least important. *Privacy* (NFR14) stood out as the highest rated quality on FI, with an FI $M = 4.41$. This item maintained a prominent position (rank 1 on FI, rank 7 on PI_{rev} with $M = 3.78$). *Authenticity* (NFR2) ranked second on FI ($M = 4.26$) and fifth on PI_{rev} ($M = 3.85$). *Protection* (NFR15) ranked third on FI ($M = 4.26$) while holding the highest position on PI_{rev} ($M = 4.30$, rank 1). *Standard* (NFR20) ranked fourth on FI ($M = 4.26$) and fourth on PI_{rev} ($M = 3.89$).

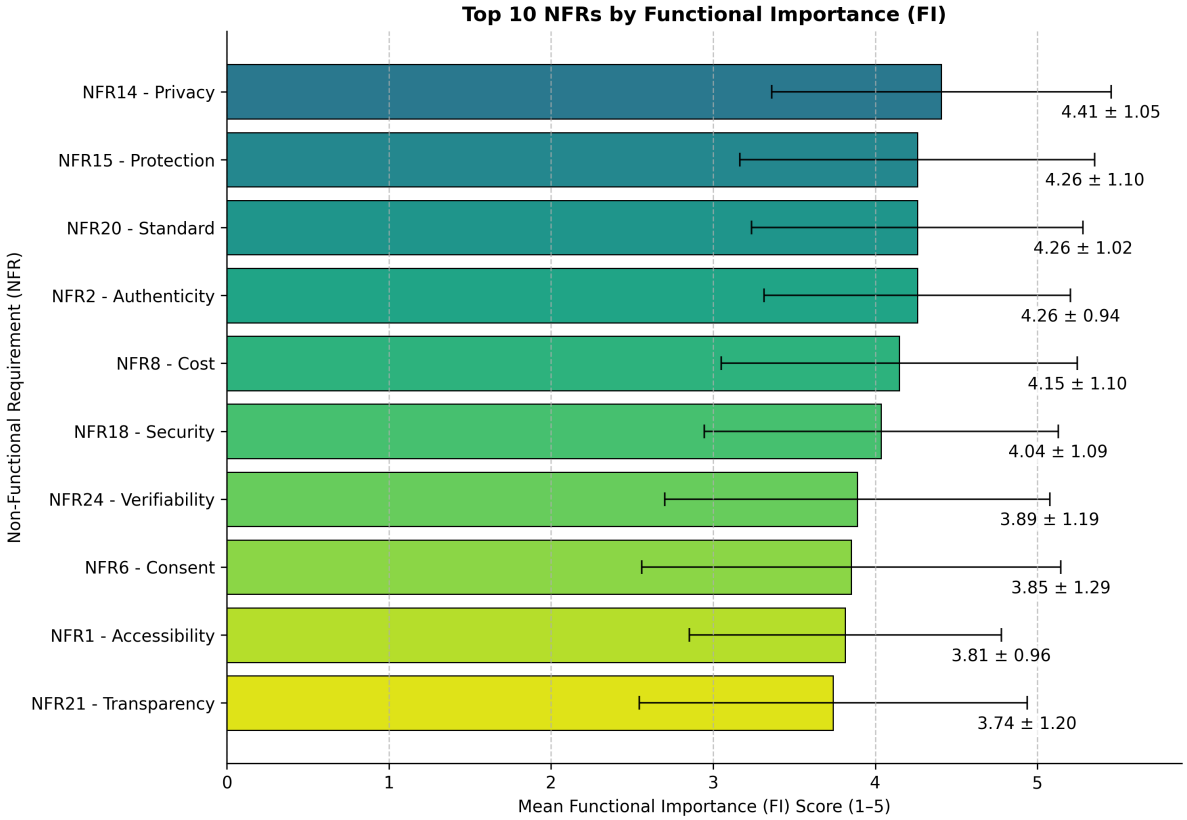

Figure 4.11: Top NFRs by FI for verifiers (mean ± SD, 1–5 scale).

On the problem based importance scale (PI_{rev}), a broadly similar set of qualities was viewed as most critical, though with some differences in ordering. *Protection* (NFR15) emerged as the highest on PI_{rev} ($M = 4.30$, rank 1), followed by *Consent* (NFR6) ($M = 4.00$, rank 2) and *Security* (NFR18) ($M = 3.96$, rank 3). *Consent* ranked substantially higher on PI_{rev} (rank 2) than on FI (rank 8). *Verifiability* (NFR24) showed the opposite pattern: it ranked seventh on FI ($M = 3.89$) but dropped to twelfth on PI_{rev} ($M = 3.19$).

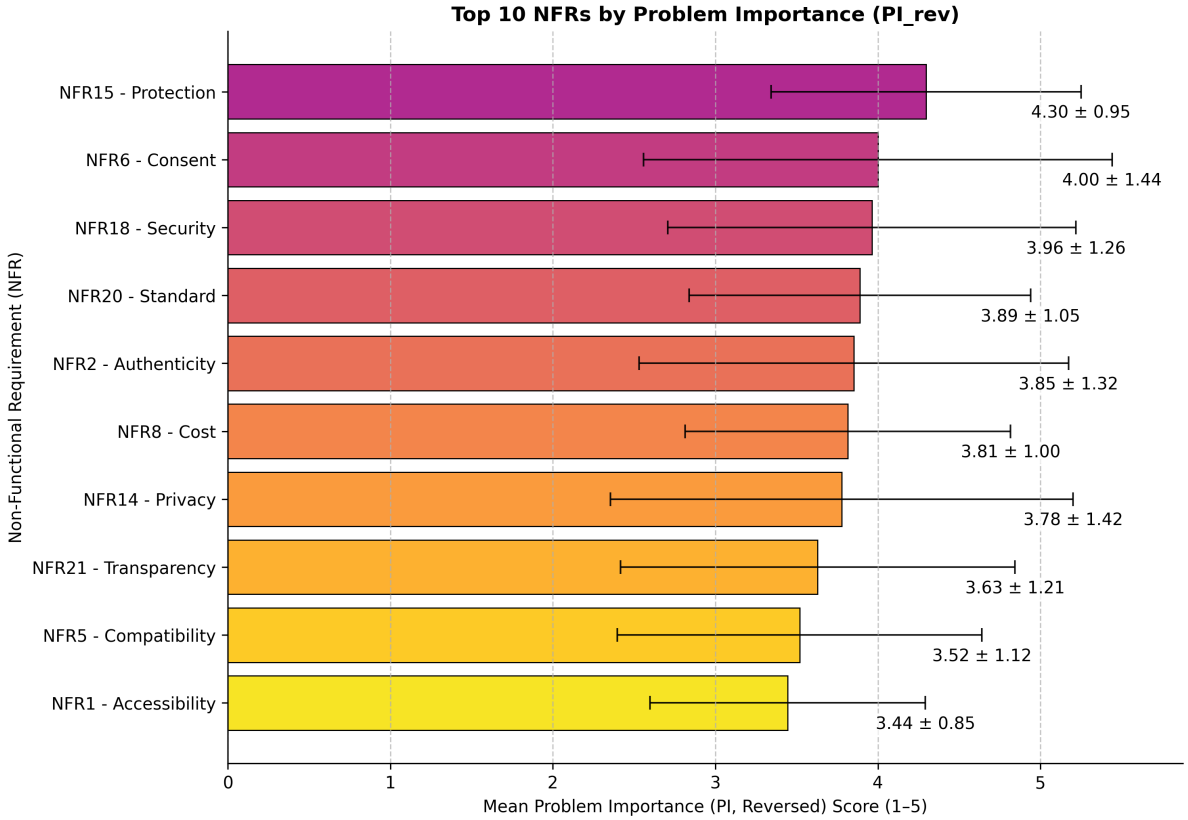

Figure 4.12: Top NFRs by $\mathrm{PI}_{\mathrm{rev}}$ for verifiers (mean ± SD, 1–5 scale).

At the lower end of the rankings, *Decentralization* (NFR9) received the lowest score on FI ($M = 3.11$, rank 13) and ranked eleventh on $\mathrm{PI}_{\mathrm{rev}}$ ($M = 3.22$). *Interoperability* (NFR11) also scored low (FI $M = 3.33$, rank 12; $\mathrm{PI}_{\mathrm{rev}}$ $M = 3.15$, rank 13, the lowest on $\mathrm{PI}_{\mathrm{rev}}$).

To provide a comprehensive overview of all 13 NFRs, several key items merited explicit attention. *Cost* (NFR8) ranked fifth on FI ($M = 4.15$) and sixth on $\mathrm{PI}_{\mathrm{rev}}$ ($M = 3.81$). *Security* (NFR18) maintained strong rankings on both scales (FI rank 6, $M = 4.04$; $\mathrm{PI}_{\mathrm{rev}}$ rank 3, $M = 3.96$). *Accessibility* (NFR1) occupied the mid upper range (FI $M = 3.81$, rank 9; $\mathrm{PI}_{\mathrm{rev}}$ $M = 3.44$, rank 10). *Transparency* (NFR21) ranked tenth on both FI ($M = 3.74$) and $\mathrm{PI}_{\mathrm{rev}}$ ($M = 3.63$). *Compatibility* (NFR5) ranked eleventh on both scales (FI $M = 3.52$; $\mathrm{PI}_{\mathrm{rev}}$ $M = 3.52$).

4.2.2.6 Friedman Rank Tests – Verifiers

The set of NFR items was treated as a repeated measures factor and analyzed using the Friedman test to identify overall differences in priorities. Kendall's W was reported as an index of agreement. When the result was significant, Dunn-Bonferroni pairwise comparisons were performed to examine differences in mean ranks, with effect sizes calculated as $r = |z|/\sqrt{n}$. NFRs were then ranked by their Friedman mean ranks, and only the five most considerable pairwise differences were highlighted. Heatmaps with stars for significance levels and color coded effect sizes were used to visualize where contrasts between NFRs

were most pronounced. This approach provided a non parametric, stakeholder specific view of NFR priorities, robust for Likert data and directly comparable between the FI and PI_{rev} blocks.

The Friedman test for verifiers showed significant differences in how NFRs were prioritized. For FI, the test yielded $\chi^2(12) = 48.87$, $p = 2.20 \times 10^{-6}$, $W = 0.151$, $n = 27$. For PI_{rev}, results were $\chi^2(12) = 35.17$, $p = 4.40 \times 10^{-4}$, $W = 0.109$, $n = 27$.

Table 4.19: Friedman mean ranks by NFR for Verifiers (FI and PI_{rev} blocks).

FI Rank	**Code**	**Name**	**FI MeanRank**	**PI_{rev} MeanRank**	**PI_{rev} Rank**
1	NFR14	Privacy	8.94	7.63	4
2	NFR15	Protection	8.37	8.87	1
3	NFR20	Standard	8.30	7.56	5
4	NFR2	Authenticity	8.06	7.41	6
5	NFR8	Cost	7.57	7.31	7
6	NFR18	Security	7.41	7.85	3
7	NFR24	Verifiability	7.02	5.46	13
8	NFR6	Consent	6.96	8.67	2
9	NFR1	Accessibility	6.57	5.85	10
10	NFR21	Transparency	6.33	6.78	8
11	NFR5	Compatibility	5.61	6.37	9
12	NFR11	Interoperability	5.54	5.69	11
13	NFR9	Decentralization	4.31	5.56	12

Note. Both blocks shown, ordered by FI rank; higher mean rank indicates higher relative priority.

Among the 78 pairwise item comparisons, post-hoc testing revealed 4 significant pairs (5.1%) in the FI block and 0 significant pairs in the PI_{rev} block.

Table 4.20: Top five strongest and weakest Bonferroni-significant pairwise contrasts by block (Verifiers).

Block	Type	Pair	Pair (names)	r	p_{adj}
FI	Strongest	NFR14 – NFR9	Privacy > Decentralization	0.841	9.79×10^{-4}
FI	Strongest	NFR15 – NFR9	Protection > Decentralization	0.736	0.0101
FI	Strongest	NFR20 – NFR9	Standard > Decentralization	0.723	0.0134
FI	Strongest	NFR2 – NFR9	Authenticity > Decentralization	0.679	0.0325
FI	Weakest	NFR2 – NFR9	Authenticity > Decentralization	0.679	0.0325
FI	Weakest	NFR20 – NFR9	Standard > Decentralization	0.723	0.0134
FI	Weakest	NFR15 – NFR9	Protection > Decentralization	0.736	0.0101
FI	Weakest	NFR14 – NFR9	Privacy > Decentralization	0.841	9.79×10^{-4}
PI_{rev}	Strongest	–	No significant pairs	–	–
PI_{rev}	Weakest	–	No significant pairs	–	–

Note. Only the five strongest and five weakest significant pairs are reported for each block. Directionality reflects the higher mean rank for the first NFR; r values are effect sizes for difference, and p_{adj} are adjusted for family-wise error.

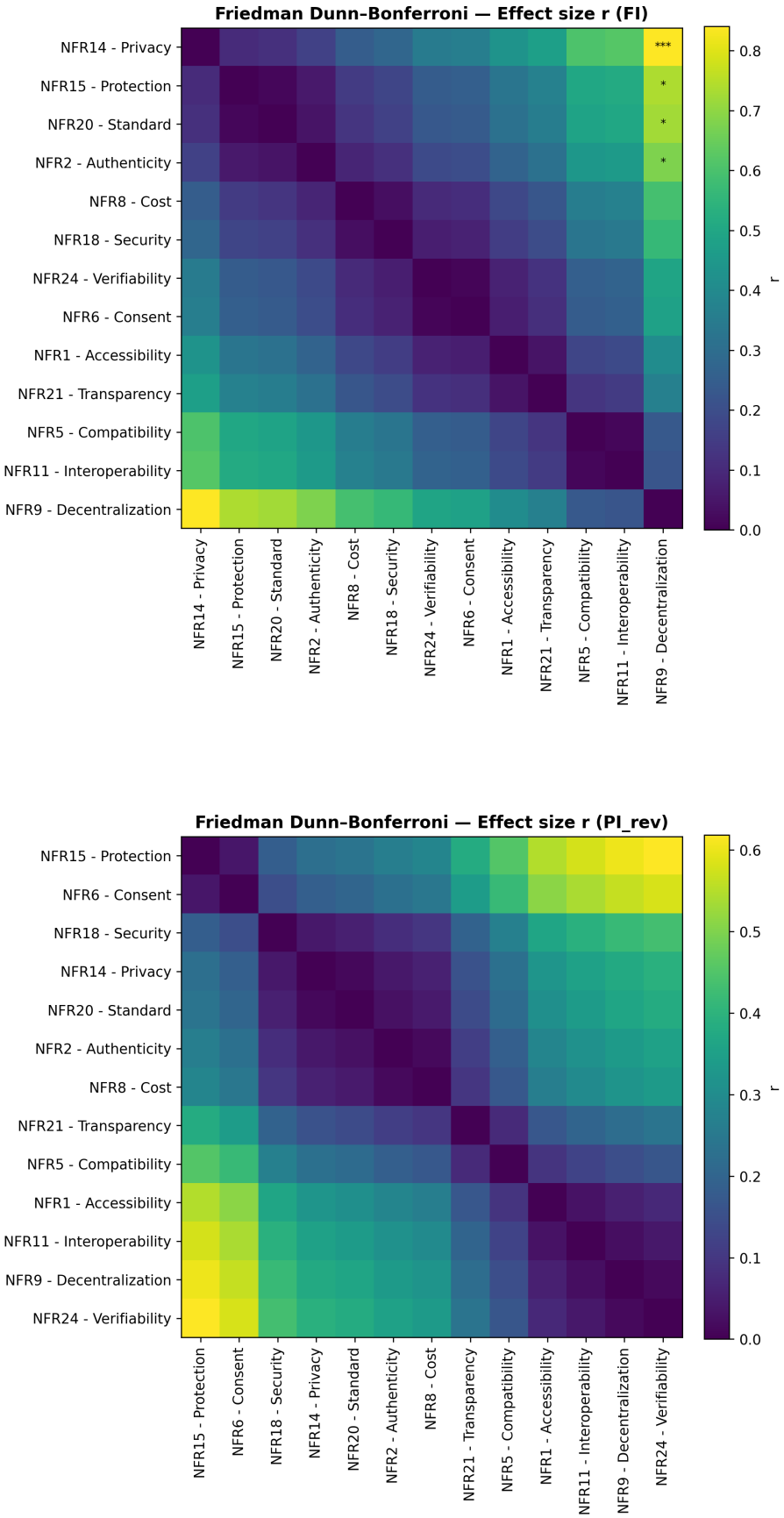

Figure 4.13: Friedman post-hoc effect-size heatmaps (r).

4.2.2.7 Prioritization matrix ($\mathbf{FI_{mean}} \times \mathbf{PI}_{rev,\mathbf{mean}}$) – Verifiers

A two dimensional scatterplot matrix was created plotting mean FI scores (x axis) against mean PI_{rev} scores (y axis) for the 13 NFRs. Quadrant thresholds were established at the within role medians (solid lines; FI $\approx$ 3.95, $PI_{rev} \approx$ 3.74), while dashed lines showed the role means for reference (FI $\approx$ 3.89, $PI_{rev} \approx$ 3.67). This divided the matrix into four zones: High FI / High PI (top right), High FI / Low PI (bottom right), Low FI / High PI (top left), and Low FI / Low PI (bottom left). Figure 4.2.2.7 illustrated the color coded quadrants.

High FI / High PI: Top priority attributes
The High FI / High PI quadrant contained 6 items, representing the core set of quality requirements that verifiers rated as both functionally critical and adequately delivered in current DI solutions. This quadrant spanned approximately FI = 4.04 to 4.41 and PI_{rev} = 3.78 to 4.30, with quadrant averages of FI $\approx$ 4.23 and $PI_{rev} \approx$ 3.93.

Table 4.21: High FI / High PI quadrant: Top-priority attributes for verifiers (n = 6)

Code	NFR Item	FI Mean	PI_{rev} Mean
NFR14	Privacy	4.41	3.78
NFR15	Protection	4.26	4.30
NFR20	Standard	4.26	3.89
NFR2	Authenticity	4.26	3.85
NFR8	Cost	4.15	3.81
NFR18	Security	4.04	3.96

High FI / Low PI quadrant: Important in principle; Lower problem salience
For verifiers, the High FI / Low PI quadrant contains one item, *Verifiability* (FI = 3.89, PI = 3.19).

Low FI / High PI quadrant: Over-delivered attributes
The Low FI / High PI quadrant contains one item, scoring below the FI median but above the PI_{rev} median: NFR6 – *Consent* (FI = 3.85, PI_{rev} = 4.00).

Low FI / Low PI quadrant: Lowest-priority attributes
The Low FI / Low PI quadrant contains 5 items, all scoring below the median thresholds on both dimensions (average FI $\approx$ 3.50, average $PI_{rev} \approx$ 3.39). These NFRs represent the lowest strategic priority for verifiers.

Table 4.22: Low FI / Low PI quadrant: Lowest-priority attributes for verifiers (n = 6)

Code	NFR Item	FI Mean	PI_{rev} Mean
NFR1	Accessibility	3.81	3.44
NFR21	Transparency	3.74	3.63
NFR5	Compatibility	3.52	3.52
NFR11	Interoperability	3.33	3.15
NFR9	Decentralization	3.11	3.22

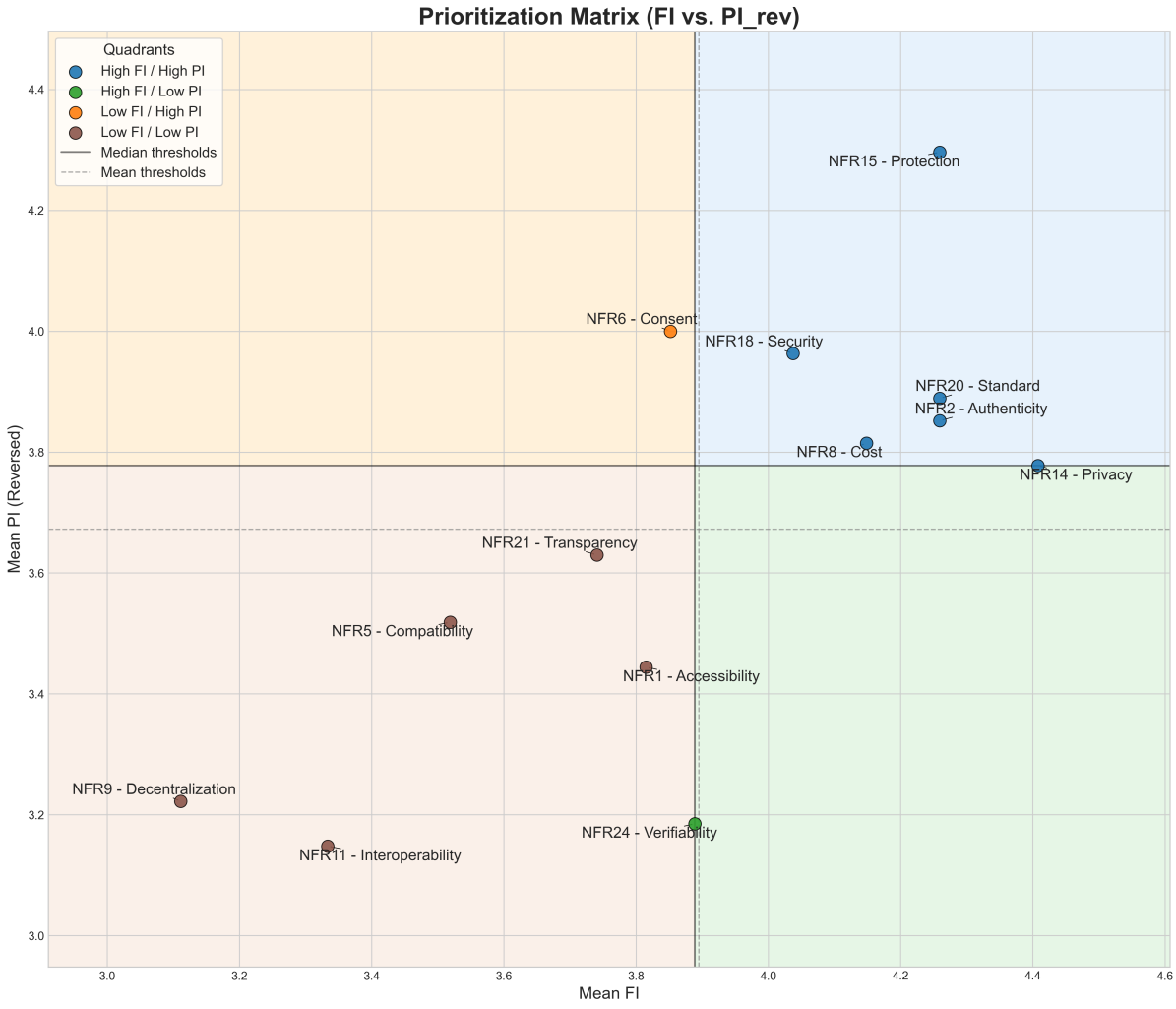

Figure 4.14: Prioritization matrix for the NFRs, displaying mean FI versus mean PI_{rev}. Shading and solid crosshairs indicate the median-based classification, while dashed crosshairs represent mean values for reference.

4.2.2.8 Group Differences Across Profession & Gender – Verifiers

To assess whether NFR priorities varied systematically by respondent demographics, non-parametric Kruskal–Wallis rank-sum tests were conducted separately for each of the 13

items in both the FI and PI_{rev} blocks, comparing ratings across gender (three groups: Female, Male, No Answer) and professional role (two groups: Manager/Executive, Other).

The Kruskal–Wallis H statistic tests the null hypothesis that all groups are drawn from the same distribution, making it appropriate for ordinal Likert-scale data with unequal group sizes. For each comparison, the test statistic H, raw p-value, ε^2 effect size (ϵ^2), and Holm-adjusted p-value ($p_{\text{adj.}}$) were computed to control the family-wise error rate within each comparison family.

Gender comparisons

Across the FI block, no items showed statistically significant gender differences after Holm correction ($p_{\text{adj.}} = 1.00$ for all 13 items). The five items with the smallest raw p-values (before correction) are displayed in Table 4.23.

Table 4.23: Kruskal-Wallis tests across gender (Female, Male, No Answer) for the FI block (Verifiers, $n = 27$).

NFR item	H	p **(raw)**	ε^2	$p_{\text{adj.}}$	**Female**	**Male**	**Non-bin.**
Compatibility (NFR5)	4.24	.120	0.093	1.00	2.78	3.87	4.00
Authenticity (NFR2)	2.97	.226	0.041	1.00	4.00	4.40	4.33
Security (NFR18)	2.86	.239	0.036	1.00	3.67	4.27	4.00
Decentralization (NFR9)	2.77	.250	0.032	1.00	2.67	3.47	2.67
Transparency (NFR21)	2.62	.270	0.026	1.00	3.33	4.00	3.67

Notes. The table lists the five items with the smallest *raw* p-values; none remained significant after Holm correction across 13 items. KW-tests across three gender groups; family-wise error controlled within the FI block using Holm adjustment over 13 items. After correction, no items were significant ($p_{\text{adj.}} = 1.00$). Across all 13 items, omnibus effect sizes were small on average (mean $\varepsilon^2 = 0.019$, range 0.0000–0.093).

Across the PI_{rev} block, no items showed significant gender differences after Holm correction ($p_{\text{adj.}} = 1.00$ for all 13 items). The five items with the smallest raw p-values are shown in Table 4.24.

Table 4.24: Kruskal–Wallis tests across gender (Female, Male, No Answer) for the PI_{rev} block (Verifiers, $n = 27$).

NFR item	H	p **(raw)**	ε^2	$p_{\text{adj.}}$	**Female**	**Male**	**Non-bin.**
Accessibility (NFR1)	3.36	.187	0.057	1.00	3.00	3.73	3.33
Interoperability (NFR11)	2.48	.289	0.020	1.00	3.00	3.40	2.33
Transparency (NFR21)	1.84	.398	0.000	1.00	3.22	3.87	3.67
Protection (NFR15)	1.62	.444	0.000	1.00	4.11	4.47	4.00
Decentralization (NFR9)	1.31	.520	0.000	1.00	3.22	3.33	2.67

Notes. The table lists the five items with the smallest *raw* p-values; none remained significant after Holm correction across 13 items. KW-tests across three gender groups; family-wise error controlled within the PI_{rev} block using Holm adjustment over 13 items. After correction, no items were significant ($p_{\text{adj.}} = 1.00$). Omnibus effects were small on average (mean $\varepsilon^2 = 0.006$, range 0.0000–0.057).

Profession comparisons

Across the FI block, no items differed significantly by professional role after Holm correction (all $p_{\text{adj.}} = 1.00$). The five items with the smallest raw p-values are shown in Table 4.25.

Table 4.25: Kruskal–Wallis tests across professional role (Manager/Executive, Other) for the FI block (Verifiers, $n = 27$).

NFR item	H	p **(raw)**	ε^2	$p_{\textbf{adj.}}$	**Manager/Exec**	**Other**
Consent (NFR6)	2.20	.138	0.063	1.00	4.19	3.20
Security (NFR18)	1.76	.185	0.040	1.00	4.25	3.40
Privacy (NFR14)	1.10	.293	0.005	1.00	4.69	4.20
Transparency (NFR21)	0.89	.346	0.000	1.00	3.88	3.20
Interoperability (NFR11)	0.67	.415	0.000	1.00	3.19	2.80

Notes. Five items with the smallest *raw* p-values; none remained significant after Holm correction across 13 items (all $p_{\text{adj.}} = 1.00$). KW-tests across two role groups; Holm correction within the FI block across 13 items (all $p_{\text{adj.}} = 1.00$). Omnibus effects were small on average (mean $\varepsilon^2 = 0.008$, range 0.0000–0.063).

Across the PI_{rev} block, no items showed significant role differences after Holm correction ($p_{\text{adj.}} = 1.00$ for all 13 items). The five items with the smallest raw p-values are presented in Table 4.26.

Table 4.26: Kruskal–Wallis tests across professional role (Manager/Executive, Other) for the PI_{rev} block (Verifiers, $n = 27$).

NFR item	H	p **(raw)**	ε^2	$p_{\textbf{adj.}}$	**Manager/Exec**	**Other**
Consent (NFR6)	3.06	.080	0.108	1.00	4.38	3.40
Authenticity (NFR2)	1.25	.264	0.013	1.00	4.00	3.00
Protection (NFR15)	0.74	.390	0.000	1.00	4.25	3.80
Interoperability (NFR11)	0.65	.422	0.000	1.00	3.13	2.60
Privacy (NFR14)	0.50	.482	0.000	1.00	3.94	3.40

Notes. Five items with the smallest *raw* p-values; none remained significant after Holm correction across 13 items. KW-tests across two role groups; family-wise error controlled within the PI_{rev} block using Holm adjustment over 13 items ($p_{\text{adj.}} = 1.00$). Mean omnibus effect size $\varepsilon^2 = 0.009$ (max 0.108 for NFR6).

SSI experience comparisons

Across the FI block, no items differed significantly by SSI experience after adjustment (all $p_{\text{adj.}} \geq .05$; range 0.44–1.00). The five items with the smallest raw p-values are shown in Table 4.27.

Table 4.27: Kruskal Wallis tests across SSI experience (Experienced vs. No experience) for the FI block (Verifiers).

NFR item	H	p **(raw)**	ε^2	$p_{\textbf{adj.}}$	**Experienced**	**No experience**
Protection (NFR15)	4.506	0.034	0.140	0.44	4.800	3.941
Transparency (NFR21)	3.273	0.070	0.091	0.84	4.200	3.471
Cost (NFR8)	3.253	0.071	0.090	0.84	4.600	3.882
Accessibility (NFR1)	2.920	0.087	0.077	0.87	4.200	3.588
Authenticity (NFR2)	2.532	0.112	0.061	1.00	4.600	4.059

Notes. Five items with the smallest *raw* p-values; none remained significant after adjustment across 24 items (all $p_{\text{adj.}}$ range 0.44–1.00).

Across the PI_{rev} block, no items showed significant differences by SSI experience after adjustment (all $p_{\text{adj.}} \geq .05$). The five items with the smallest raw p-values are presented in Table 4.28.

Table 4.28: Kruskal Wallis tests across SSI experience (Experienced vs. No Experience) for the PI*rev* block (Verifiers).

NFR item	H	p **(raw)**	ε^2	$p_{\textbf{adj.}}$	**Experienced**	**No experience**
Standard (NFR20)	2.565	0.109	0.063	1.00	4.300	3.647
Transparency (NFR21)	2.123	0.145	0.045	1.00	4.000	3.412
Protection (NFR15)	2.023	0.155	0.041	1.00	4.600	4.118
Verifiability (NFR24)	1.428	0.232	0.017	1.00	3.500	3.000
Accessibility (NFR1)	1.379	0.240	0.015	1.00	3.200	3.588

Notes. Five items with the smallest *raw* p-values.

Post-hoc pairwise comparisons

Following the hierarchical testing procedure, pairwise Dunn tests with Bonferroni-adjusted p-values and rank-biserial effect sizes were pre-specified only for items whose omnibus Kruskal–Wallis test reached significance after Holm correction at $\alpha = .05$. None of the 78 omnibus tests (13 items $\times$ 2 blocks $\times$ 3 grouping variables) met this criterion after family-wise error control; therefore, no post-hoc comparisons were conducted.

4.2.2.9 Item-Level Descriptives & Rankings – Issuers

In the following section, descriptive statistics for each NFR were calculated on both the FI and PI_{rev} scales, including the mean, median, and SD of the ratings. These statistics summarize the central tendency and variability of perceived importance for each quality attribute. To facilitate interpretation, each item was also ranked within the set of 12 NFRs based on its mean score on each scale. A rank of 1 indicates the highest mean importance (most important), and the highest rank number corresponds to the lowest importance relative to other items. Table 4.29 and Figures 4.15 and 4.16 present the item-level descriptive results and rankings side by side for the FI and PI_{rev} measures.

Table 4.29: Item–level descriptive statistics and importance rankings for the 12 NFR items (SQRI) for issuers.

	FI				**PI_{rev}**			
NFR Item	**Mean**	**Median**	**SD**	**Rank**	**Mean**	**Median**	**SD**	**Rank**
NFR15 – Protection	4.73	5.0	0.61	1	4.73	5.0	0.77	2
NFR18 – Security	4.70	5.0	0.57	2	4.84	5.0	0.44	1
NFR14 – Privacy	4.57	5.0	0.65	3	4.05	4.0	1.15	6
NFR2 – Authenticity	4.51	5.0	0.69	4	4.30	4.0	0.85	4
NFR20 – Standard	4.41	5.0	0.72	5	3.00	3.0	1.41	11
NFR8 – Cost	4.30	5.0	0.88	6	3.92	4.0	1.06	7
NFR11 – Interoperability	4.24	4.0	0.76	7	3.22	3.0	1.13	10
NFR24 – Verifiability	4.08	4.0	1.12	8	4.43	5.0	0.90	3
NFR6 – Consent	3.95	4.0	1.25	9	3.81	4.0	1.43	8
NFR5 – Compatibility	3.73	4.0	0.99	10	2.86	3.0	1.06	12
NFR21 – Transparency	3.41	4.0	1.46	11	4.16	5.0	1.07	5
NFR9 – Decentralization	3.35	3.0	1.18	12	3.24	3.0	1.04	9

Note. For each NFR, the table reports the mean, median, and standard deviation (SD) of the FI ratings and PI_{rev} ratings, along with the item's rank on each scale (where 1 indicates the highest importance).

From these results, several clear patterns emerged regarding which NFRs issuers found most and least important. *Protection* (NFR15) stood out as the highest rated quality on FI, with an FI $M = 4.73$. This item maintained its top position across both scales (PI_{rev} $M = 4.73$, rank 2). *Security* (NFR18) held the highest position on PI_{rev} ($M = 4.84$, rank 1) while ranking second on FI ($M = 4.70$, rank 2). *Privacy* (NFR14) ranked third on FI ($M = 4.57$) and sixth on PI_{rev} ($M = 4.05$). *Authenticity* (NFR2) ranked fourth on both scales (FI $M = 4.51$, PI_{rev} $M = 4.30$).

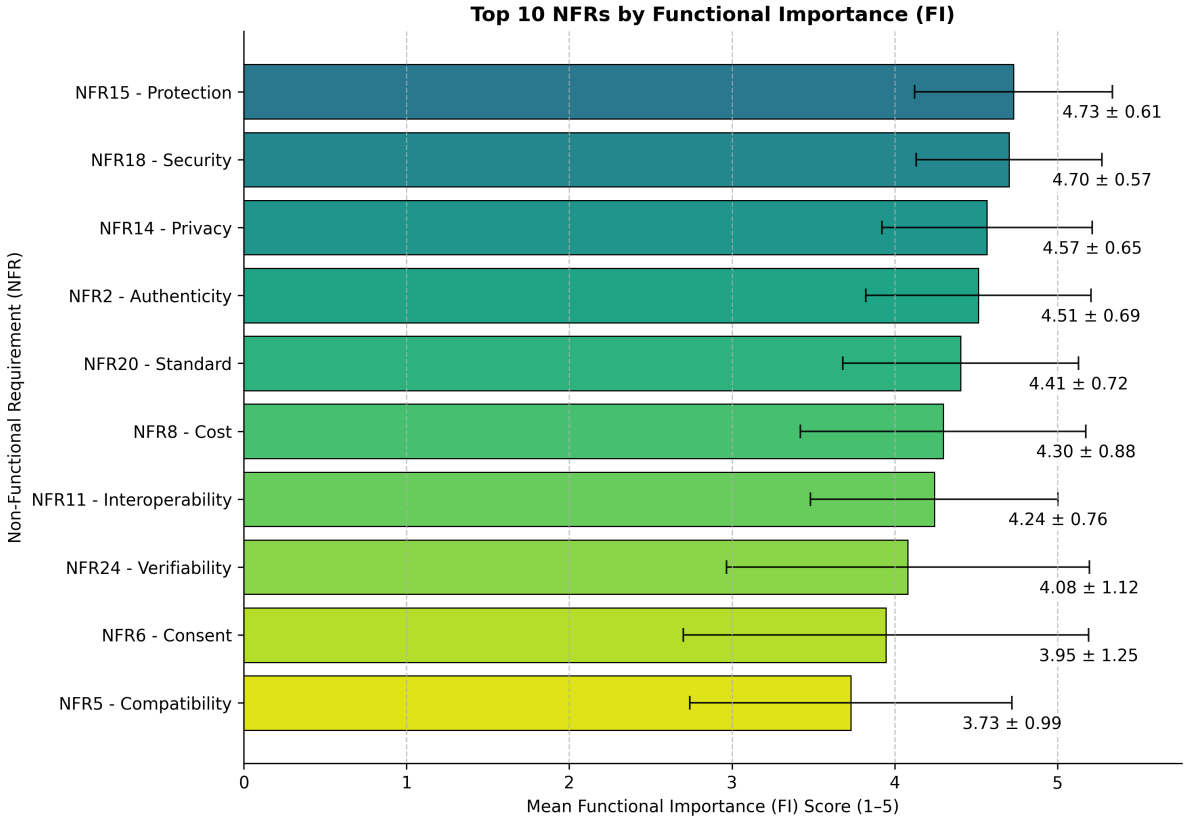

Figure 4.15: Top NFRs by FI for issuers (mean ± SD, 1–5 scale).

On the $\mathrm{PI}_{\mathrm{rev}}$ scale, a broadly similar set of qualities was viewed as most critical, though with some differences in ordering. *Security* (NFR18) emerged as the highest on $\mathrm{PI}_{\mathrm{rev}}$ ($M = 4.84$, rank 1), followed by *Protection* (NFR15) ($M = 4.73$, rank 2) and *Verifiability* (NFR24) ($M = 4.43$, rank 3). *Verifiability* ranked substantially higher on $\mathrm{PI}_{\mathrm{rev}}$ (rank 3) than on FI (rank 8). *Transparency* (NFR21) showed the opposite pattern: it ranked relatively low on FI ($M = 3.41$, rank 11) but rose to fifth on $\mathrm{PI}_{\mathrm{rev}}$ ($M = 4.16$, rank 5).

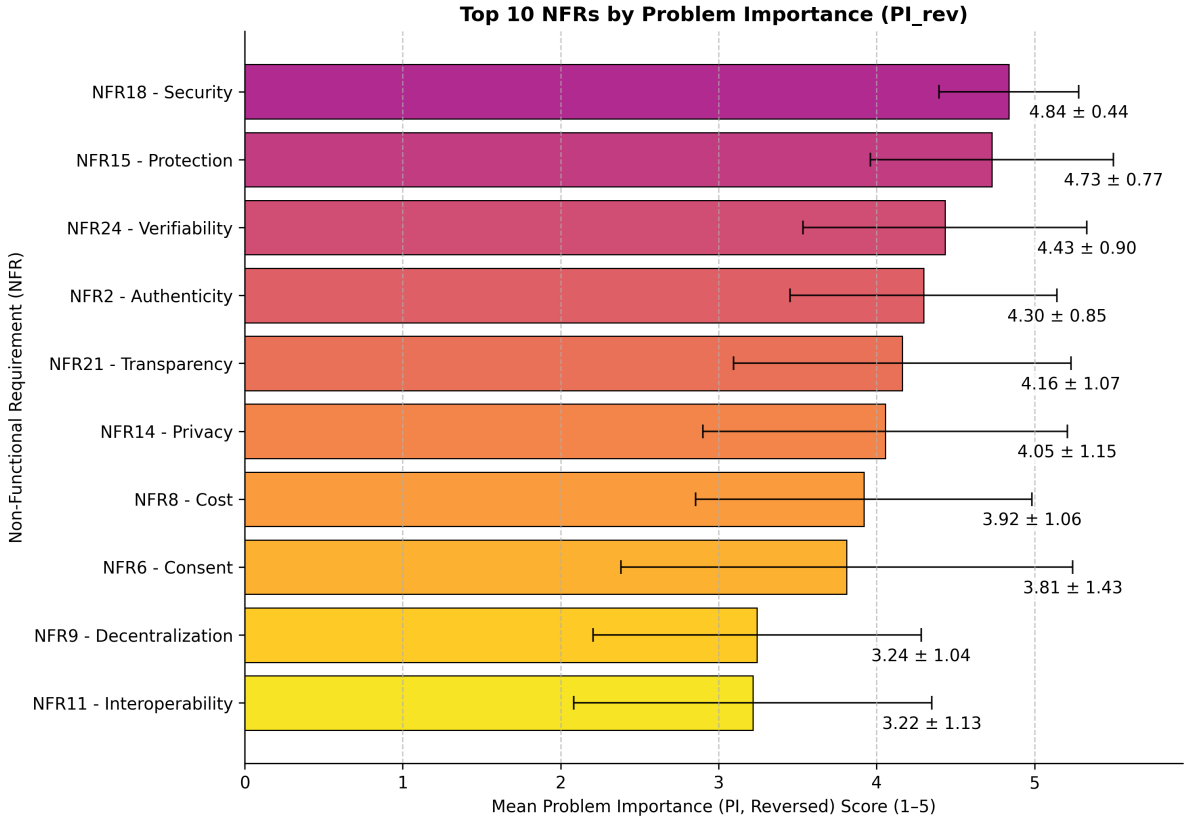

Figure 4.16: Top NFRs by PI_{rev} for issuers (mean ± SD, 1–5 scale).

At the lower end of the rankings, *Decentralization* (NFR9) received the lowest score on FI ($M = 3.35$, rank 12) and ranked ninth on PI_{rev} ($M = 3.24$). *Compatibility* (NFR5) also scored low (FI $M = 3.73$, rank 10; PI_{rev} $M = 2.86$, rank 12, the lowest on PI_{rev}).

To provide a comprehensive overview of all 12 NFRs, several key items merited explicit attention. *Standard* (NFR20) ranked fifth on FI ($M = 4.41$) but dropped to eleventh on PI_{rev} ($M = 3.00$). *Cost* (NFR8) ranked sixth on FI ($M = 4.30$) and seventh on PI_{rev} ($M = 3.92$). *Interoperability* (NFR11) maintained mid tier rankings on both scales (FI rank 7, $M = 4.24$; PI_{rev} rank 10, $M = 3.22$). *Consent* (NFR6) occupied the upper mid range (FI $M = 3.95$, rank 9; PI_{rev} $M = 3.81$, rank 8).

4.2.2.10 Friedman Rank Tests – Issuers

The set of NFR items was treated as a repeated measures factor and analyzed using the Friedman test to identify overall differences in priorities. Kendall's W was reported as an index of agreement. When the result was significant, Dunn Bonferroni pairwise comparisons were performed to examine differences in mean ranks, with effect sizes calculated as $r = |z|/\sqrt{n}$. NFRs were then ranked by their Friedman mean ranks, and only the five most considerable pairwise differences were highlighted. Heatmaps with stars for significance levels and color coded effect sizes visualized where contrasts between NFRs were most pronounced. This approach provided a non parametric, stakeholder specific view of

NFR priorities, robust for Likert data and directly comparable between the FI and PI_{rev} blocks.

The Friedman test for issuers showed significant differences in how NFRs were prioritized. For FI, the test yielded $\chi^2(11) = 85.81$, $p = 1.10 \times 10^{-13}$, $W = 0.211$, $n = 37$. For PI_{rev}, results were $\chi^2(11) = 134.85$, $p = 1.82 \times 10^{-23}$, $W = 0.331$, $n = 37$.

Table 4.30 presented the Friedman mean ranks for all 12 NFRs in the FI and PI_{rev} blocks for issuers. Each mean rank was based on assigning a rank from 1 to 12 per respondent and averaging these ranks across participants (1 = lowest priority, 12 = highest priority). *Protection* and *Security* occupied the top two positions across blocks; *Verifiability* and *Transparency* moved up under problem framing.

Table 4.30: Friedman mean ranks by NFR for Issuers (FI and PI_{rev} blocks).

FI Rank	**Code**	**Name**	**FI MeanRank**	**PI_{rev} MeanRank**	**PI_{rev} Rank**
1	NFR15	Protection	8.41	9.12	2
2	NFR18	Security	8.26	9.41	1
3	NFR14	Privacy	7.73	6.93	6
4	NFR2	Authenticity	7.34	7.30	4
5	NFR20	Standard	7.03	4.34	11
6	NFR8	Cost	6.88	6.41	8
7	NFR11	Interoperability	6.47	4.54	10
8	NFR24	Verifiability	6.38	7.73	3
9	NFR6	Consent	6.04	6.65	7
10	NFR5	Compatibility	4.91	3.62	12
11	NFR21	Transparency	4.51	7.27	5
12	NFR9	Decentralization	4.05	4.69	9

Note. Ordered by FI rank; higher mean rank indicates higher relative priority.

Post-hoc pairwise comparisons revealed significant differences in 13 of the $\binom{12}{2} = 66$ NFR pairs (19.7%) for the FI block and in 19 of 66 pairs (28.8%) for the PI_{rev} block. Table 4.31 lists the five strongest and five weakest Bonferroni-significant contrasts for each block (direction indicates the higher mean rank).

Table 4.31: Top five strongest and weakest Bonferroni-significant pairwise contrasts by block (Issuers).

Block	**Type**	**NFR Pair**	**Pair (names)**	r	p_{adj}
FI	Strongest	NFR15 – NFR9	Protection > Decentralization	0.853	1.38×10^{-5}
FI	Strongest	NFR9 – NFR18	Security > Decentralization	0.824	3.53×10^{-5}
FI	Strongest	NFR15 – NFR21	Protection > Transparency	0.763	0.0002
FI	Strongest	NFR21 – NFR18	Security > Transparency	0.734	0.0005
FI	Strongest	NFR14 – NFR9	Privacy > Decentralization	0.721	0.0008
FI	Weakest	NFR8 – NFR9	Cost > Decentralization	0.554	0.0498
FI	Weakest	NFR2 – NFR21	Authenticity > Transparency	0.554	0.0498
FI	Weakest	NFR5 – NFR14	Privacy > Compatibility	0.554	0.0498
FI	Weakest	NFR20 – NFR9	Standard > Decentralization	0.583	0.0258
FI	Weakest	NFR14 – NFR21	Privacy > Transparency	0.631	0.0082
PI_{rev}	Strongest	NFR5 – NFR18	Security > Compatibility	1.134	3.44×10^{-10}
PI_{rev}	Strongest	NFR5 – NFR15	Protection > Compatibility	1.079	3.53×10^{-9}
PI_{rev}	Strongest	NFR20 – NFR18	Security > Standard	0.994	9.85×10^{-8}
PI_{rev}	Strongest	NFR11 – NFR18	Security > Interoperability	0.954	4.29×10^{-7}
PI_{rev}	Strongest	NFR20 – NFR15	Protection > Standard	0.938	7.60×10^{-7}
PI_{rev}	Weakest	NFR20 – NFR21	Transparency > Standard	0.575	0.0309
PI_{rev}	Weakest	NFR2 – NFR20	Authenticity > Standard	0.580	0.0274
PI_{rev}	Weakest	NFR8 – NFR18	Security > Cost	0.588	0.0228
PI_{rev}	Weakest	NFR5 – NFR6	Consent > Compatibility	0.594	0.0201
PI_{rev}	Weakest	NFR24 – NFR9	Verifiability > Decentralization	0.596	0.0189

Note. Only the five strongest and five weakest significant pairs are reported for each block. Directionality reflects the higher mean rank for the first NFR; r values are effect sizes for difference, and p_{adj} are adjusted for family-wise error.

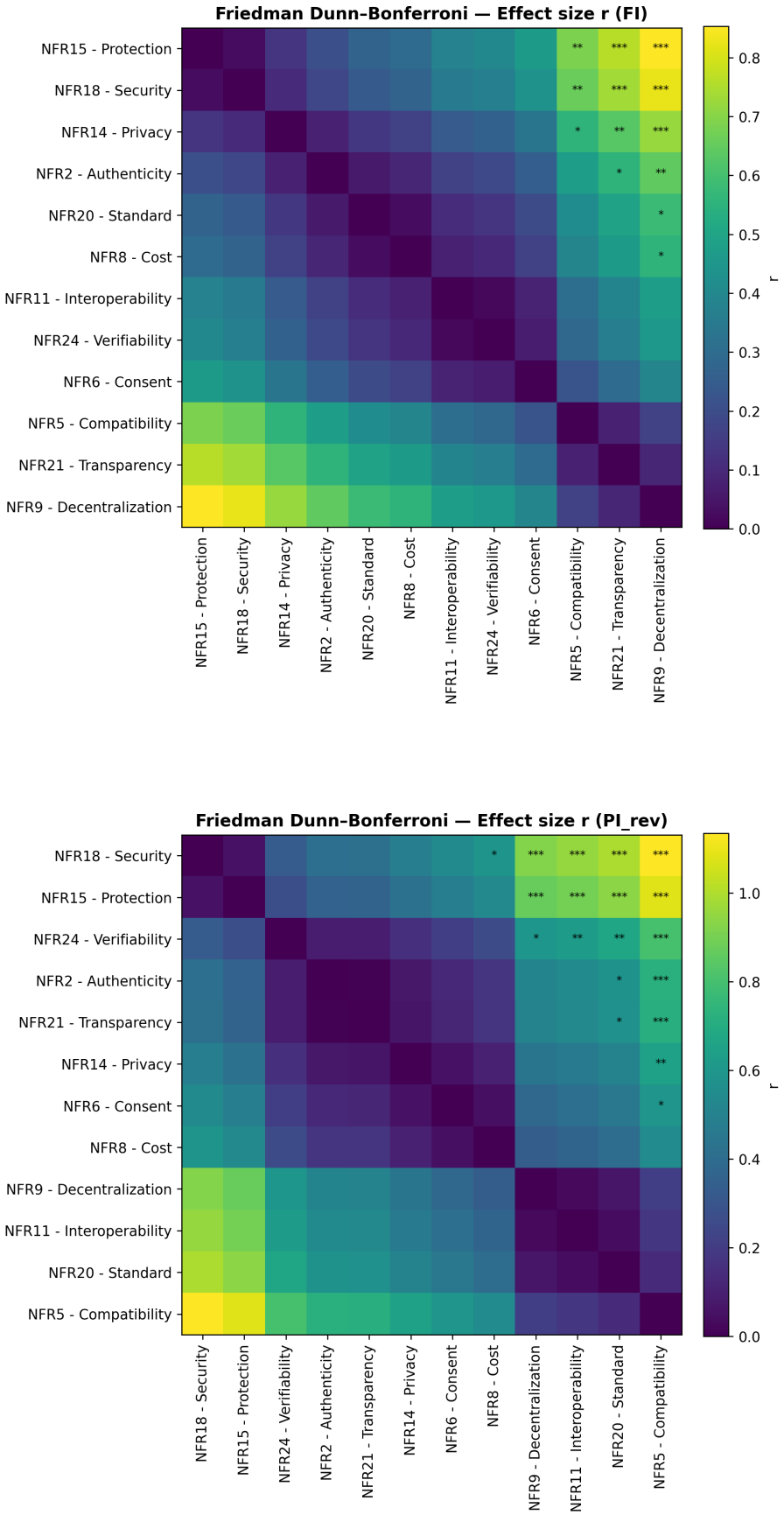

Figure 4.17: Friedman post-hoc effect-size heatmaps for issuers.

4.2.2.11 Prioritization matrix ($\mathbf{FI_{mean}} \times \mathbf{PI}_{rev,\mathbf{mean}}$) – Issuers

A two dimensional scatterplot matrix was created to combine the dimensions of importance and problem salience into a cohesive prioritization framework, with mean FI scores represented on the x axis and mean PI_{rev} scores on the y axis. Quadrant thresholds were established at the medians of the 12 item means (FI median = 4.20, PI_{rev} median = 4.00), which is more robust to skew and ceiling effects than mean splits. These medians divided the matrix into four strategic zones: High FI / High PI (top right), High FI / Low PI (top left), Low FI / High PI (bottom right), and Low FI / Low PI (bottom left). Figure 4.18 showed the median crosshairs (solid) and the mean reference lines (dashed).

High FI / High PI: Top-priority attributes
The High FI / High PI quadrant contains four items, representing the core set that issuers rate as both functionally critical and comparatively high in problem salience. This set spans FI = 4.51–4.73 and PI_{rev} = 4.05–4.84, with quadrant averages FI $\approx$ 4.62 and PI_{rev} $\approx$ 4.48.

Table 4.32: High FI / High PI quadrant: Top-priority attributes for issuers ($n = 4$)

Code	NFR Item	FI Mean	$\mathbf{PI_{rev}}$ Mean
NFR18	Security	4.70	4.84
NFR15	Protection	4.73	4.73
NFR2	Authenticity	4.51	4.30
NFR14	Privacy	4.57	4.05

High FI / Low PI quadrant: Important in principle; Lower problem salience
The High FI / Low PI quadrant comprises two items (quadrant averages FI $\approx$ 4.42, PI_{rev} $\approx$ 3.47), indicating qualities that issuers value strongly in principle but that fall below the median on perceived problem salience.

Table 4.33: High FI / Low PI quadrant: Important in principle; lower problem salience ($n = 2$)

Code	NFR Item	FI Mean	$\mathbf{PI_{rev}}$ Mean
NFR8	Cost	4.43	3.94
NFR20	Standard	4.41	3.00

Low FI / High PI quadrant: Lower importance; Higher problem salience
The Low FI / High PI quadrant contains two items, reflecting attributes that rise in priority under problem framing despite sitting below the FI median.

Table 4.34: Low FI / High PI quadrant: Lower importance; higher problem salience ($n = 2$)

Code	NFR Item	FI Mean	$\mathbf{PI_{rev}}$ Mean
NFR24	Verifiability	4.08	4.43
NFR21	Transparency	3.41	4.16

Low FI / Low PI quadrant: Lowest-priority attributes

The Low FI / Low PI quadrant contains four items (quadrant averages FI ≈ 3.78, $\mathrm{PI_{rev}}$ ≈ 3.28), representing the lowest strategic priority under the median classification.

Table 4.35: Low FI / Low PI quadrant: Lowest-priority attributes for issuers ($n = 4$)

Code	NFR Item	FI Mean	$\mathbf{PI_{rev}}$ Mean
NFR11	Interoperability	4.11	3.22
NFR6	Consent	3.95	3.81
NFR5	Compatibility	3.72	2.86
NFR9	Decentralization	3.34	3.23

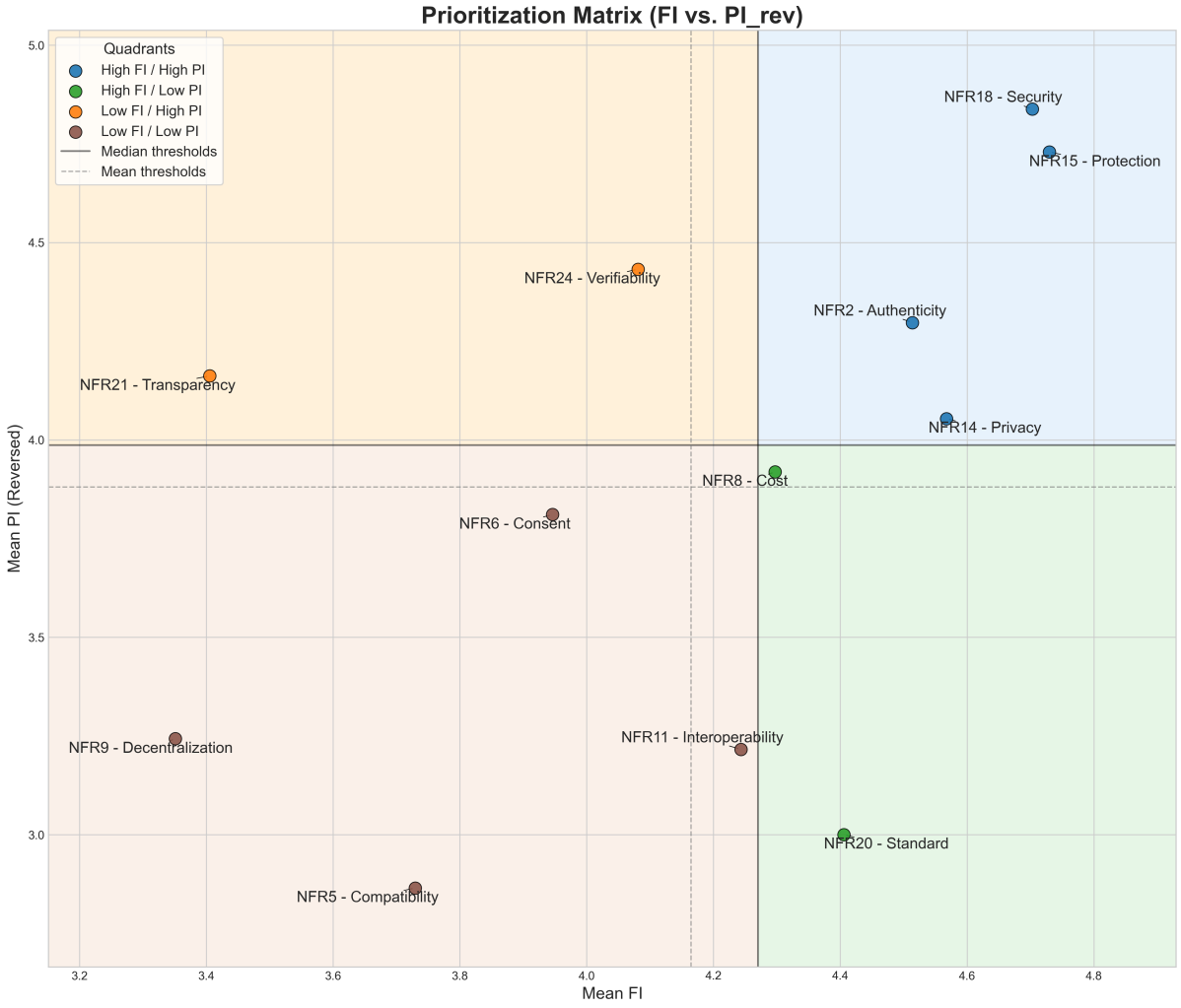

Figure 4.18: Prioritization matrix for the NFRs, displaying mean FI versus mean $\mathrm{PI}_{\mathrm{rev}}$. Shading and solid crosshairs indicate the median-based classification, while dashed crosshairs represent mean values for reference.

4.2.2.12 Group Differences Across Profession & Gender – Issuers

To assess whether NFR priorities varied systematically by respondent demographics among issuers, non parametric Kruskal Wallis rank sum tests were conducted separately for each of the 12 items in both the FI and PI_{rev} blocks, comparing ratings across gender (three groups: Female, Male, No Answer) and professional role (three groups: Manager/Executive, Professional/Academic occupation, Other).

The Kruskal Wallis H statistic tested the null hypothesis that all groups were drawn from the same distribution, making it appropriate for ordinal Likert scale data with unequal group sizes. For each comparison, the test statistic H, raw p value, epsilon squared effect size (ϵ^2, a rank based measure of association strength ranging from 0 to 1), and Holm adjusted p value ($p_{adj.}$) were computed to control the family wise error rate within each comparison family. Complete test results, group wise descriptive means, and post hoc pairwise comparisons are provided in an Excel file across 12 sheets (see Appendix B): separate Kruskal Wallis results, means tables, and post hoc tables for FI and PI_{rev} by gender and role.

Gender comparisons
Across the FI block for issuers, no items showed statistically significant gender differences after Holm correction ($p_{adj.} \geq .05$ for all 12 items, range 0.11–1.00). The five items with the smallest raw p-values (before correction) are displayed in Table 4.36.

Table 4.36: Kruskal–Wallis tests across gender (Female, Male, No Answer) for the FI block (Issuers, $n = 37$).

NFR item	H	p **(raw)**	ε^2	$p_{adj.}$	**Female**	**Male**	**Non-bin.**
Verifiability (NFR24)	9.33	.009	0.215	0.11	4.00	4.32	1.67
Transparency (NFR21)	2.41	.300	0.012	1.00	3.67	3.52	2.00
Compatibility (NFR5)	1.82	.402	0.000	1.00	4.33	3.71	3.33
Privacy (NFR14)	1.69	.430	0.000	1.00	4.33	4.55	5.00
Authenticity (NFR2)	1.55	.462	0.000	1.00	4.00	4.55	4.67

Notes. The table lists the five items with the smallest *raw* p-values; none remained significant after Holm correction across 12 items. KW-tests across three gender groups; family-wise error controlled within the FI block using Holm adjustment over 12 items. After correction, no items were significant ($p_{adj.} \geq 0.05$; range 0.11–1.00). Across all 12 items, omnibus effect sizes were small on average (mean $\varepsilon^2 = 0.019$, range 0.0000–0.215).

Across the PI_{rev} block for issuers, no items showed significant gender differences after Holm correction ($p_{adj.} = 1.00$ for all 12 items). The five items with the smallest raw p-values are shown in Table 4.37.

Table 4.37: Kruskal–Wallis tests across gender (Female, Male, No Answer) for the PI_{rev} block (Issuers, $n = 37$).

NFR item	H	p **(raw)**	ε^2	$p_{\text{adj.}}$	**Female**	**Male**	**Non-bin.**
Standard (NFR20)	4.82	.090	0.083	1.00	4.67	2.84	3.00
Interoperability (NFR11)	3.39	.184	0.041	1.00	2.00	3.32	3.33
Privacy (NFR14)	3.26	.196	0.037	1.00	4.33	3.94	5.00
Verifiability (NFR24)	2.81	.245	0.024	1.00	3.33	4.52	4.67
Consent (NFR6)	2.34	.310	0.010	1.00	4.67	3.65	4.67

Notes. The table lists the five items with the smallest *raw* p-values; none remained significant after Holm correction across 12 items. KW-tests across three gender groups; family-wise error controlled within the PI_{rev} block using Holm adjustment over 12 items. After correction, no items were significant (all $p_{\text{adj.}} = 1.00$). Omnibus effects were minor on average (mean $\varepsilon^2 = 0.016$, range 0.0000–0.083).

Profession comparisons

Across the FI block, no items differed significantly by professional role among issuers after Holm correction ($p_{\text{adj.}} \geq .05$ for all 12 items, range 0.11–1.00). The five items with the smallest raw p-values are shown in Table 4.38.

Table 4.38: Kruskal–Wallis tests across professional role (Manager/Executive, Professional/Academic, Other) for the FI block (Issuers, $n = 37$).

NFR item	H	p **(raw)**	ε^2	$p_{\text{adj.}}$	**Manager**	**Prof/Acad**	**Other**
Compatibility (NFR5)	9.47	.009	0.299	0.11	3.94	2.75	4.50
Standard (NFR20)	4.42	.110	0.097	1.00	4.19	4.75	4.75
Decentralization (NFR9)	4.31	.116	0.092	1.00	3.75	3.12	2.50
Interoperability (NFR11)	3.88	.143	0.075	1.00	4.06	4.62	4.50
Security (NFR18)	3.07	.215	0.043	1.00	4.81	4.62	4.25

Notes. Five items with the smallest *raw* p-values; none remained significant after Holm correction across 12 items. KW-tests across three role groups; Holm correction within the FI block across 12 items. No items were significant after adjustment ($p_{\text{adj.}}$ range 0.11–1.00). Mean $\varepsilon^2 = 0.055$.

Across the PI_{rev} block for issuers, no items showed significant role differences after Holm correction ($p_{\text{adj.}} \geq .05$ for all 12 items, range 0.41–1.00). The five items with the smallest raw p-values are presented in Table 4.39.

Table 4.39: Kruskal–Wallis tests across professional role (Manager/Executive, Professional/Academic, Other) for the PI_{rev} block (Issuers, $n = 37$).

NFR item	H	p **(raw)**	ε^2	$p_{\text{adj.}}$	**Manager**	**Prof/Acad**	**Other**
Transparency (NFR21)	6.74	.034	0.190	0.41	4.31	4.25	2.75
Cost (NFR8)	5.25	.072	0.130	0.80	4.06	4.25	2.75
Standard (NFR20)	1.83	.401	0.000	1.00	2.50	3.25	3.25
Protection (NFR15)	1.63	.442	0.000	1.00	4.50	4.88	5.00
Compatibility (NFR5)	1.39	.499	0.000	1.00	2.88	2.88	3.50

Notes. Five items with the smallest *raw* p-values; none remained significant after Holm correction across 12 items. KW-tests across three role groups; family-wise error controlled within the PI_{rev} block using Holm adjustment over 12 items. No items significant ($p_{\text{adj.}}$ range 0.41–1.00). Mean $\varepsilon^2 = 0.027$.

SSI experience comparisons

Across the FI block, no items differed significantly by SSI experience after adjustment (all $p_{\text{adj.}} = 1.00$). The five items with the smallest raw p-values are shown in Table 4.40. Omnibus effects were very small on average ($\bar{\varepsilon}^2 = 0.004$, range 0.000 to 0.021, maximum for *Authenticity*, NFR2).

Table 4.40: Kruskal–Wallis tests across SSI experience (Experienced vs. No Experience) for the FI block (Issuers).

NFR item	H	p **(raw)**	ε^2	$p_{\text{adj.}}$	**No SSI exp.**	**SSI exp.**
Authenticity (NFR2)	2.63	.105	0.021	1.00	4.40	4.78
Security (NFR18)	2.25	.133	0.017	1.00	4.28	4.67
Interoperability (NFR11)	2.06	.152	0.015	1.00	4.05	4.50
Transparency (NFR21)	1.74	.187	0.011	1.00	3.82	4.33
Portability (NFR13)	1.59	.207	0.009	1.00	3.92	4.33

Notes. Five items with the smallest *raw* p-values; none remained significant after adjustment across 12 items (all $p_{\text{adj.}} = 1.00$). KW tests across two groups (Experienced vs. No Experience); adjustment within the block across 12 items.

Across the PI_{rev} block, no items showed significant differences by SSI experience after adjustment (all $p_{\text{adj.}} = 1.00$). The five items with the smallest raw p-values are presented in Table 4.41. Omnibus effects were also very small on average ($\bar{\varepsilon}^2 = 0.003$, range 0.000 to 0.017, maximum for *Protection*, NFR15).

Table 4.41: Kruskal–Wallis tests across SSI experience (Experienced vs. No Experience) for the PI*rev* block (Issuers).

NFR item	H	p **(raw)**	ε^2	$p_{\text{adj.}}$	**No SSI exp.**	**SSI exp.**
Protection (NFR15)	2.11	.147	0.017	1.00	4.44	4.78
Consent (NFR6)	1.85	.173	0.014	1.00	4.33	4.67
Authenticity (NFR2)	1.76	.185	0.012	1.00	4.28	4.61
Privacy (NFR14)	1.52	.217	0.009	1.00	4.11	4.44
Transparency (NFR21)	1.38	.241	0.007	1.00	3.89	4.22

Notes. Five items with the smallest *raw* p-values; none remained significant after adjustment across 12 items (all $p_{\text{adj.}} = 1.00$). KW tests across two groups (Experienced vs. No Experience); adjustment within the block across 12 items.

Post-hoc pairwise comparisons

Following the hierarchical testing procedure, pairwise Dunn tests with Bonferroni-adjusted p-values and rank-biserial effect sizes were pre-specified only for items whose omnibus Kruskal–Wallis test remained significant after Holm correction at $\alpha = .05$. None of the 72 omnibus tests (12 items $\times$ 2 blocks $\times$ 3 grouping variables) met this criterion after family-wise error control; therefore, no post-hoc comparisons were conducted.

4.2.3 BWS Analysis

The analysis of BWS data provides a complementary way to assess priorities among quality requirements. In the BWS tasks, participants viewed sets of requirements and selected the most and least important within each set. To analyze these choices, the script first calculated standardized best-worst scores for each requirement by subtracting the number of times an item was chosen as least important from the number of times it was chosen as most important, and dividing by the total times the item appeared in choice sets. These scores range from -1 to $+1$ and offer an easy-to-understand summary of preferences across all participants. Standardized count scores are reported for descriptive transparency. However, they are not used as the primary analysis because count-based methods only reflect choice frequencies and ignore context, for example, whether an item was selected as "worst" in a very strong set or a very weak one [15].

However, this simple count method treats all choices equally. It ignores which items were shown together in each set, as well as the possibility that some requirements might appear more frequently or in different contexts than others [15]. To address this, aggregated choice data were analyzed using a MNL-based exploded-logit (rank-ordered) specification estimated via maximum likelihood, explicitly accounting for the choice set structure [15, 61]. For identification, one requirement was set to $\beta = 0$ as the reference, which is standard in logit models [85]. In this formulation, each best-worst task is represented as two sequential choices: respondents first select the most important item from the complete set and then the least important item from the remaining options. In the exploded-logit model, this is implemented as two "best of the available alternatives" choices (first from the complete set, then from the reduced set), which is mathematically equivalent to placing the observed "worst" item at the bottom of the ranking [61, 85]. This specification yields utility coefficients (beta, β) and associated standard errors that reflect both the frequency and the context of best and worst choices, enabling statistical comparisons between requirements and providing information on the uncertainty in the estimated scores [15]. Although some NFRs exhibit higher or even positive net scores, the exploded-logit model additionally accounts for the specific choice context and competing items in each set. As a result, model-based coefficients can diverge slightly from simple count patterns, and small differences among lower-ranked NFRs should be interpreted as statistically uncertain rather than substantively meaningful.

Results of the BWS analysis are presented in both tabular and graphical formats. A table lists each requirement, its importance score from the model, its standard error, and its standardized best-worst score. Items are ranked by importance, making it easy to compare priorities. Bar charts display each requirement's score in descending order for each stakeholder group, while additional plots show the cumulative importance shares and highlight gaps between top and lower-ranked items. These visualizations help to interpret the pattern of priorities revealed by participants' best-worst choices.

The use of both count-based and model-based methods in this study provides a transparent and robust summary of how stakeholders prioritize quality requirements, with clear descriptive and statistical information. No formal hypotheses were tested. The analysis focused on assessing and describing the observed preference structure in the data.

4.2.3.1 BWS Results – Identity Holders (Users)

As a methodological complement to the direct rating approach (SQRI), a subset of participants ($N = 78$) completed a Best Worst Scaling (BWS) discrete choice experiment, in which each respondent evaluated multiple small sets of NFRs and selected the most important ("best") and least important ("worst") item in each set. This forced choice format mitigates common biases in Likert scale data, such as response set, central tendency, and acquiescence, by requiring explicit trade offs between attributes rather than permitting uniform high or low ratings across all items. The BWS design generated repeated pairwise comparisons across all 24 NFRs, with each item exposed 234 times across all choice tasks (exposure count uniform for all items). Aggregated choice data were analyzed using an MNL-based exploded-logit (rank-ordered) specification estimated via maximum likelihood, yielding a utility coefficient (beta, β) for each NFR that quantifies its relative importance. In this model, higher β values indicate a greater propensity for an NFR to be chosen as "best" rather than "worst" within its choice sets (relative to the reference NFR), so positive β denotes above-average importance and negative β denotes below-average importance. Complete BWS model results, including beta coefficients, standard errors, z-statistics, p-values, importance shares, and raw best/worst choice frequencies, are reported in the thesis' repository (see Appendix B).

The model converged successfully, producing a full set of beta estimates with standard errors ranging from 0.15 to 0.17. Beta coefficients ranged from $\beta = 0.753$ (*Security*, NFR18, rank 1) to $\beta = -0.706$ (*Recoverability*, NFR16, rank 24), with a M of $\beta = -0.062$ and a median of $\beta = -0.184$. Ten of the 24 items (41.7%) achieved statistical significance at $p < .05$: four items with positive β values (indicating above average importance) and six items with negative β values (indicating below average importance). Beta coefficients were transformed into importance shares (percentage of total utility), which summed to 100% across all items and facilitated intuitive interpretation of relative priorities. Importance shares ranged from 8.66% (*Security*) to 2.01% (*Recoverability*), with the top five items accounting for 36.5% of total importance and the top ten items accounting for 58.4%.

Table 4.42: Complete Best-Worst Scaling results for all 24 NFRs.

Rank	NFR	β	SE	p	Share (%)	Best	Worst	Net	Norm
1	NFR18 – Security	0.753	0.162	$< .001$	8.66	143	10	+133	0.57
2	NFR15 – Protection	0.718	0.162	$< .001$	8.36	140	11	+129	0.55
3	NFR21 – Transparency	0.613	0.160	$< .001$	7.53	129	27	+102	0.44
4	NFR7 – Control	0.491	0.161	$< .01$	6.67	119	17	+102	0.44
5	NFR6 – Consent	0.263	0.166	$> .05$	5.30	102	16	+86	0.37
6	NFR5 – Compatibility	0.167	0.150	$> .05$	4.82	27	107	-80	-0.34
7	NFR8 – Cost	0.067	0.160	$> .05$	4.36	30	99	-69	-0.29
8	NFR22 – Usability	0.063	0.159	$> .05$	4.35	14	122	-108	-0.46
9	NFR14 – Privacy	0.057	0.163	$> .05$	4.32	94	20	+74	0.32
10	NFR9 – Decentralization	0.000	–	–	4.08	78	46	+32	0.14
11	NFR13 – Portability	-0.005	0.150	$> .05$	4.06	22	100	-78	-0.33
12	NFR17 – Representation	-0.165	0.148	$> .05$	3.46	15	116	-101	-0.43
13	NFR12 – Persistence	-0.202	0.154	$> .05$	3.33	37	72	-35	-0.15
14	NFR23 – User Experience	-0.210	0.150	$> .05$	3.31	20	109	-89	-0.38
15	NFR10 – Existence	-0.226	0.152	$> .05$	3.25	13	99	-86	-0.37
16	NFR4 – Availability	-0.248	0.155	$> .05$	3.18	67	43	+24	0.10
17	NFR3 – Autonomy	-0.254	0.165	$> .05$	3.16	44	59	-15	-0.06
18	NFR11 – Interoperability	-0.295	0.152	$> .05$	3.04	36	77	-41	-0.18
19	NFR19 – Single Source	-0.381	0.165	$< .05$	2.79	48	53	-5	-0.02
20	NFR24 – Verifiability	-0.437	0.165	$< .01$	2.63	49	33	+16	0.07
21	NFR20 – Standard	-0.466	0.164	$< .01$	2.56	14	80	-66	-0.28
22	NFR2 – Authenticity	-0.520	0.169	$< .01$	2.43	64	24	+40	0.17
23	NFR1 – Accessibility	-0.553	0.169	$< .01$	2.35	52	35	+17	0.07
24	NFR16 – Recoverability	-0.706	0.173	$< .001$	2.01	47	29	+18	0.08

Notes. Model estimates: β = utility coefficient, SE = standard error, p = significance threshold ($< .001$, $< .01$, $< .05$, or $> .05$), Share = importance share (%); Count metrics: Best/Worst = observed choice frequencies, Net = Best − Worst, Norm = standardized count score (Net/Exposure, range −1 to +1). All items had exposure = 234.

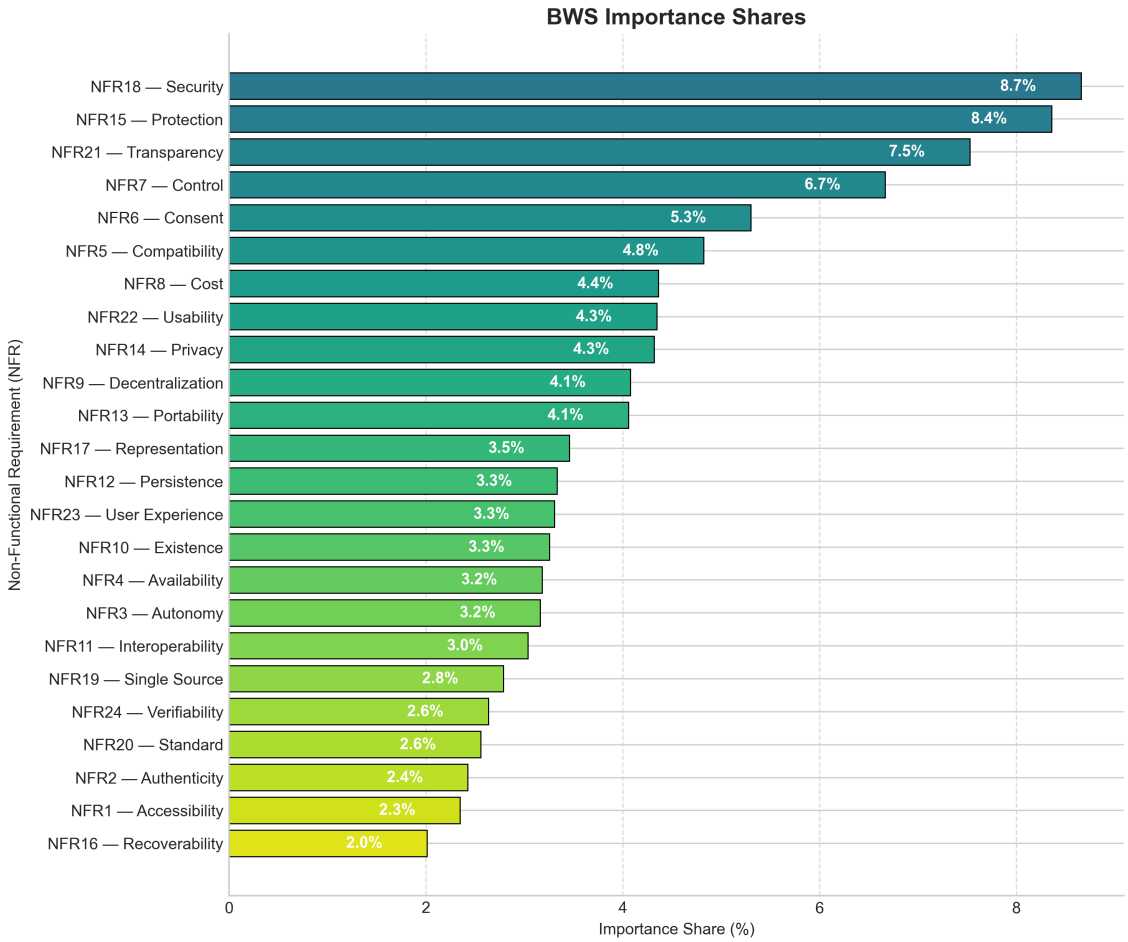

Figure 4.19: BWS importance shares ranked by model estimates. Importance shares sum to 100% across all items. Colors gradient from high (dark teal) to low (yellow) priority.

Table 4.42 and Figure 4.19 presented the complete BWS ranking of all 24 NFRs. *Security* (NFR18) emerged as the highest priority (rank 1, share = 8.66%, net = +133), followed by *Protection* (NFR15) (rank 2, share = 8.36%), *Transparency* (NFR21) (rank 3, share = 7.53%), *Control* (NFR7) (rank 4, share = 6.67%), and *Consent* (NFR6) (rank 5, share = 5.30%). These top five NFRs accounted for 36.5% of total importance, confirming the dominance of security and control attributes observed in the SQRI results. The bottom tier comprised *Recoverability* (NFR16) (rank 24, $\beta = -0.706$, share = 2.01%), *Accessibility* (NFR1) (rank 23), and *Authenticity* (NFR2) (rank 22), with importance shares below 2.5%. Because the exploded-logit model estimates utilities conditional on the specific choice sets in which NFRs appear and relative to the reference item (NFR9 – *Decentralization*), the model-based coefficients do not correspond linearly to the simple Best/Worst counts. As a result, an NFR can have a high positive Net score but still receive a comparatively low or negative β if its best choices mainly occur in easier sets. In contrast, an NFR with a negative Net can obtain a mid-level β when it frequently competes against strong NFRs.

Notably, several mid ranked items exhibited negative net scores (more "worst" than "best" choices) yet achieved middle tier rankings in the model. For example, *Usability* (NFR22)

ranked 8th (share = 4.35%) despite having the most negative net score of all items (net = −108). This occurred because the model estimated utility conditional on choice set composition: *Usability* was often selected as "worst" in sets containing even lower priority alternatives, yielding a positive beta coefficient ($\beta = 0.063$). The implications of these methodological differences and the convergence of BWS and SQRI rankings are discussed in Section 5.

4.2.3.2 Alignment between SQRI and BWS – Identity Holders (Users)

To examine alignment between rating-based and choice-based prioritization among verifiers, SQRI FI ratings and BWS importance scores were standardized to z-scores and plotted together. The axes of the prioritization matrix are divided by within-role medians, indicated by solid lines, while means are displayed with dashed lines for reference. Medians were selected for splitting because initial assumption checks showed pronounced ceiling effects and asymmetry in the 1–5 Likert-scale FI items, indicating that the median provides greater robustness to outliers and tied values than the mean. Figure 4.20 illustrates this prioritization matrix, with each point representing an *NFR* item, labeled by code and assigned a quadrant color: High–High (QI, blue), Low–High (QII, orange), Low–Low (QIII, brown), and High–Low (QIV, green).

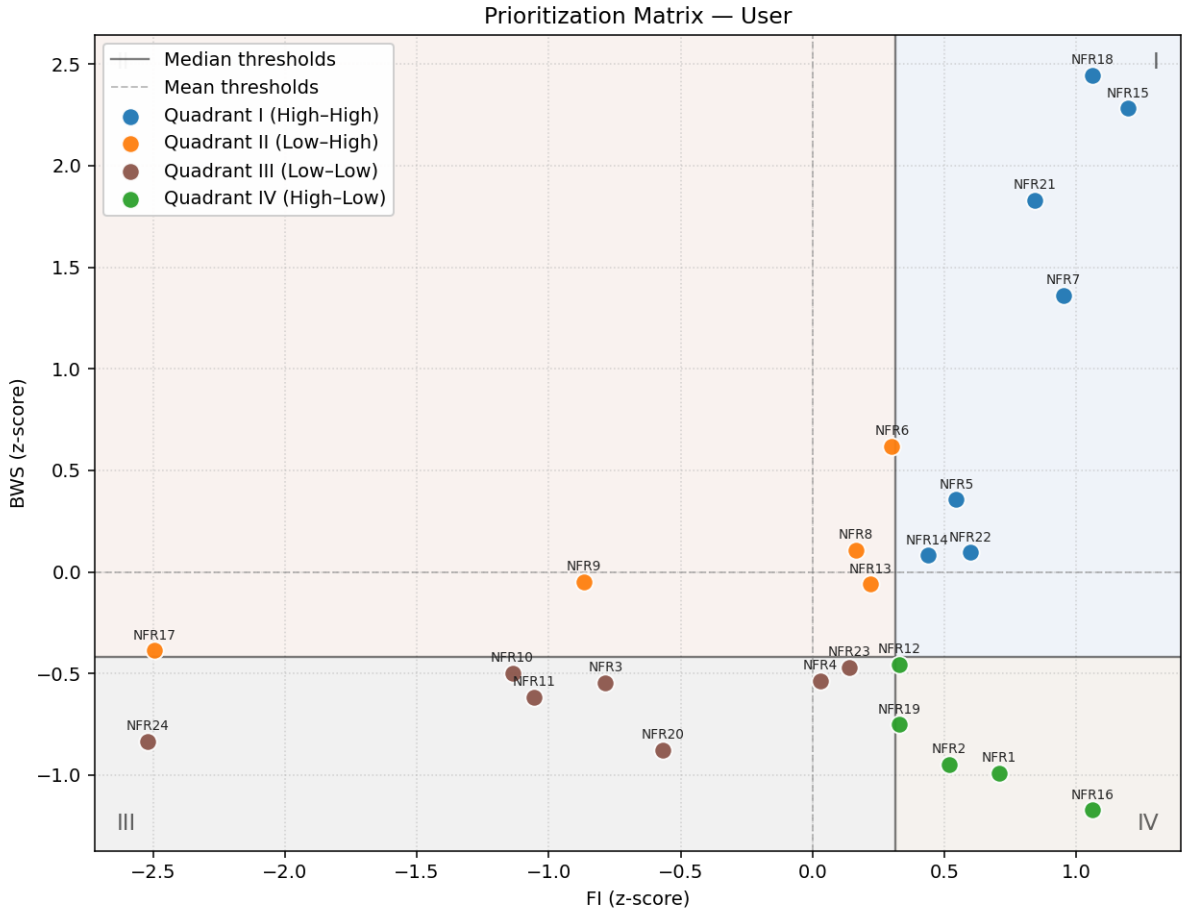

Figure 4.20: FI and BWS are z-scores; solid lines indicate median thresholds (classification), dashed lines indicate means (reference). Colors denote quadrants: I (High–High), II (Low–High), III (Low–Low), IV (High–Low).

The two approaches show a moderate positive correlation (Pearson's $r = 0.458$, $p = 0.025$; Spearman's $\rho = 0.373$, $p = 0.073$). This suggests that while there is some agreement between methods, each also captures unique aspects of NFR prioritization. In the standardized prioritization matrix, where both FI and BWS scores are expressed as z-scores and separated at within-role medians, seven NFRs fall in **Quadrant I (High–High)**: *Security* (NFR18), *Protection* (NFR15), *Transparency* (NFR21), *Control* (NFR7), *Compatibility* (NFR5), *Privacy* (NFR14), and *Usability* (NFR22). These items consistently rank above the median on both measures and form a core consensus set. *Consent* (NFR6) lies on the vertical median, indicating a borderline status when the two approaches are considered together.

Quadrant IV (High FI, Low BWS) includes *Recoverability* (NFR16), *Accessibility* (NFR1), *Authenticity* (NFR2), *Single-Source* (NFR19), and *Persistence* (NFR12), which have high Likert ratings but lower importance in choice-based tasks. In contrast, **Quadrant II (Low FI, High BWS)** includes *Consent* (NFR6), *Cost* (NFR8), *Portability* (NFR13), *Decentralization* (NFR9), and *Representation* (NFR17), which are rated as more important in BWS than on the Likert scale. Several items are close to the median so that small changes could shift them into different quadrants. Because FI scores show ceiling effects

and asymmetry, the median split is a robust classification approach. The mean lines serve only as a reference for sensitivity. The standardized matrix confirms areas of consensus (*NFR18*, *NFR15*, *NFR21*, *NFR7*) and highlights the differences between the two prioritization methods. These findings are discussed further in Section 5.

4.2.3.3 BWS Results – Verifiers

A subset of verifier participants ($N = 27$) completed a Best–Worst Scaling (BWS) discrete choice experiment, evaluating 13 NFRs with each item presented 72 times across all respondents.

The model converged successfully with beta coefficients ranging from $\beta = 0.387$ (*Authenticity*, NFR2, rank 1) to $\beta = -0.850$ (*Transparency*, NFR21, rank 13), $M = 0.009$. Only one item achieved statistical significance at $p < .05$: NFR21 (*Transparency*) with a negative β value. Importance shares ranged from 10.65% (*Authenticity*) to 3.09% (*Transparency*), with the top five items accounting for 50.4% and the top ten for 86.3% of total importance.

Table 4.43: Complete Best-Worst Scaling results for all 13 NFRs (Verifiers, $N = 27$).

Rank	**NFR**	$\boldsymbol{\beta}$	**SE**	p	**Share (%)**	**Best**	**Worst**	**Net**	**Norm**
1	NFR2 – Authenticity	0.387	0.348	$> .05$	10.65	41	8	+33	0.46
2	NFR8 – Cost	0.383	0.236	$> .05$	10.61	29	24	+5	0.07
3	NFR6 – Consent	0.375	0.284	$> .05$	10.52	30	22	+8	0.11
4	NFR14 – Privacy	0.312	0.277	$> .05$	9.88	30	22	+8	0.11
5	NFR11 – Interoperability	0.190	0.237	$> .05$	8.74	25	23	+2	0.03
6	NFR15 – Protection	0.153	0.317	$> .05$	8.43	33	15	+18	0.25
7	NFR9 – Decentralization	0.000	–	–	7.23	7	40	-33	-0.46
8	NFR24 – Verifiability	−0.022	0.367	$> .05$	7.08	36	12	+24	0.33
9	NFR1 – Accessibility	−0.030	0.352	$> .05$	7.02	11	45	-34	-0.47
10	NFR5 – Compatibility	−0.169	0.322	$> .05$	6.11	16	23	-7	-0.10
11	NFR20 – Standard	−0.285	0.352	$> .05$	5.44	12	46	-34	-0.47
12	NFR18 – Security	−0.331	0.345	$> .05$	5.19	28	12	+16	0.22
13	NFR21 – Transparency	−0.850	0.370	$< .05$	3.09	14	20	-6	-0.08

Notes. Model estimates: β = utility coefficient, SE = standard error, p = significance threshold ($< .001$, $< .01$, $< .05$, or $> .05$), Share = importance share (%); Count metrics: Best/Worst = observed choice frequencies, Net = Best − Worst, Norm = standardized count score (Net/Exposure). All items had equal exposure of 72 presentations across respondents.

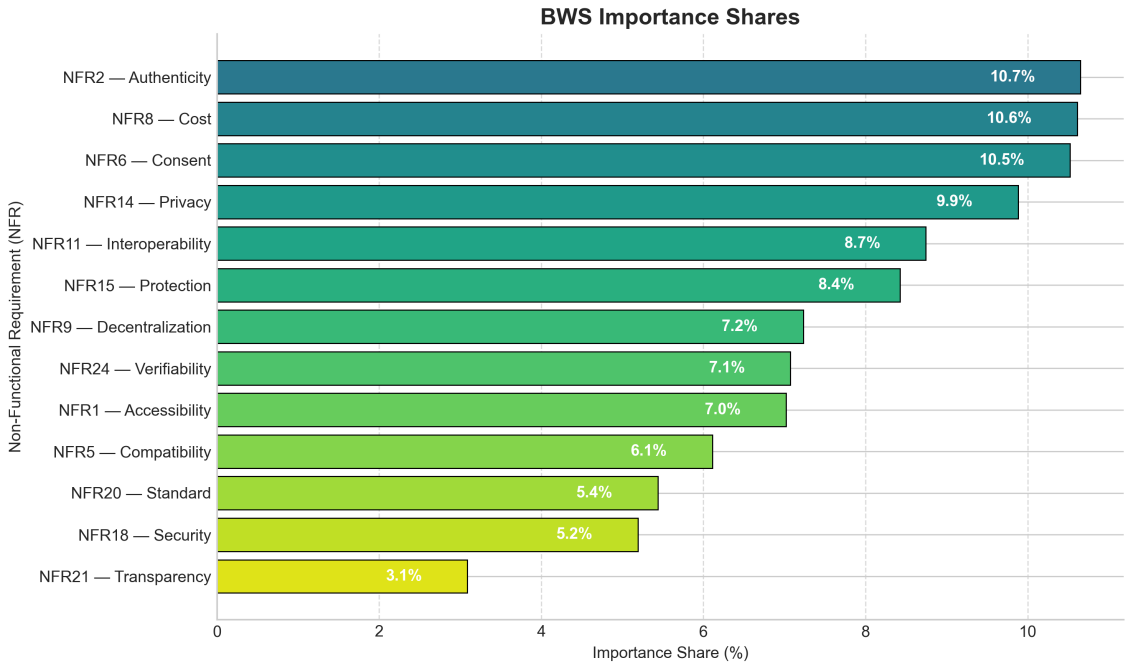

Figure 4.21: BWS importance shares ranked by model estimates. Importance shares sum to 100% across all items. Colors gradient from high (dark teal) to low (yellow) priority.

Table 4.43 and Figure 4.21 presents the complete BWS ranking of all 13 *NFRs* for verifiers. *Authenticity* (NFR2) emerges as the highest priority (rank 1, share = 10.65%, net = +33), followed by *Cost* (NFR8) (rank 2, share = 10.61%), *Consent* (NFR6) (rank 3, share = 10.52%), *Privacy* (NFR14) (rank 4, share = 9.88%), and *Interoperability* (NFR11) (rank 5, share = 8.74%). These top five *NFRs* account for 50.4% of total importance. Notably, six items exhibit positive β coefficients, indicating they were chosen as "best" more often than "worst." The bottom tier comprises *Transparency* (NFR21) (rank 13, $\beta = -0.850$, share = 3.09%, the only statistically significant item at $p < .05$), *Security* (NFR18) (rank 12), and *Standard* (NFR20) (rank 11), with importance shares below 5.5%. The relatively flat distribution across items and limited statistical significance suggest verifiers may perceive multiple NFRs as comparably important, in contrast to the stronger differentiation observed among issuers.

4.2.3.4 Alignment between SQRI and BWS – Verifiers

To examine alignment between rating-based and choice-based prioritization among verifiers, SQRI FI ratings and BWS importance scores were standardized to z-scores and plotted together. The axes of the prioritization matrix are divided by within-role medians (solid lines), while means are shown with dashed lines for reference. Medians were selected for splitting because assumption checks indicated ceiling effects and asymmetry in the 1–5 Likert FI items, for which median thresholds provide greater robustness to ties and non-normality. Figure 4.22 shows the matrix with NFR codes only and quadrant colors: High–High (QI, blue), Low–High (QII, orange), Low–Low (QIII, brown), and High–Low (QIV, green).

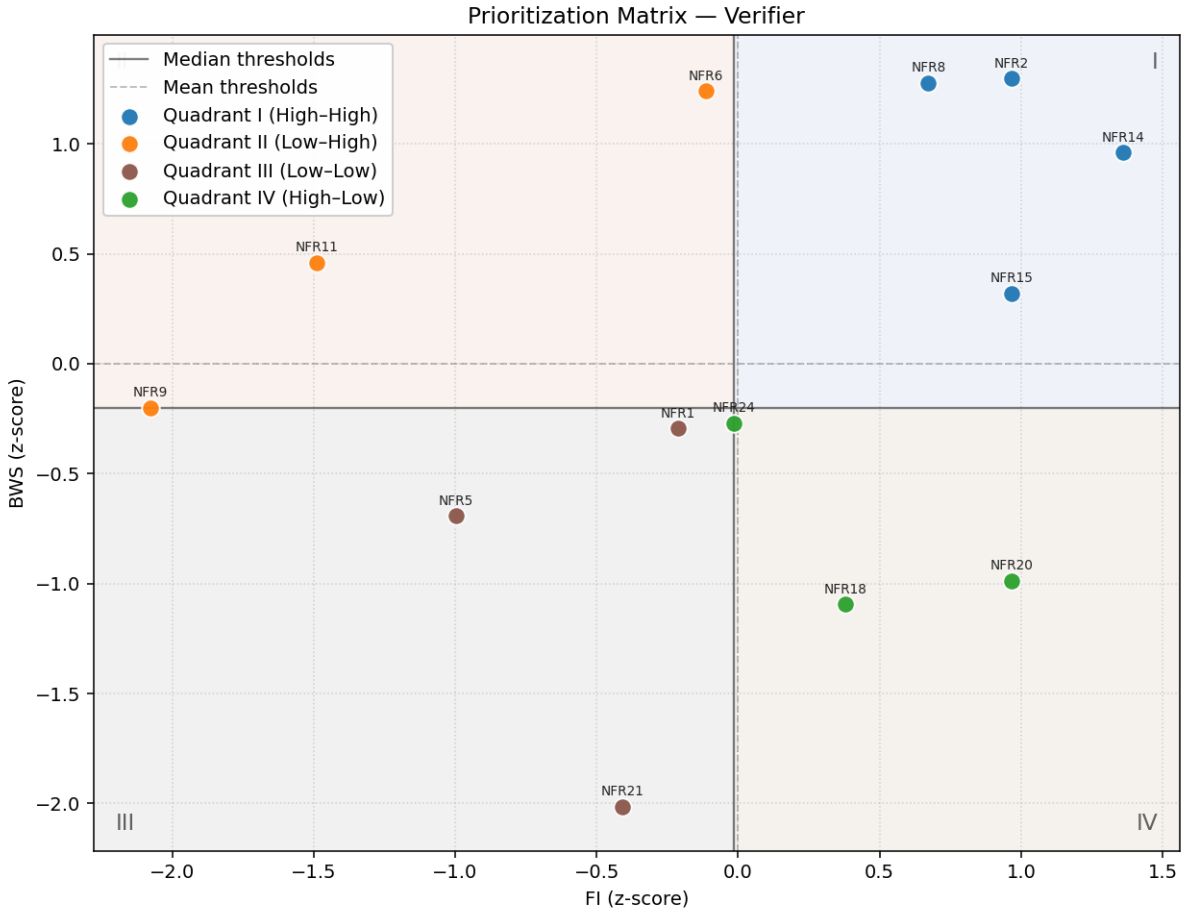

Figure 4.22: FI and BWS are z-scores; solid lines indicate median thresholds (classification), dashed lines indicate means (reference). Colors denote quadrants: I (High–High), II (Low–High), III (Low–Low), IV (High–Low).

The two approaches show a weak, non-significant positive correlation (Pearson's $r = 0.272$, $p = 0.369$; Spearman's $\rho = 0.304$, $p = 0.313$), indicating limited convergence and appreciable method-specific variance. In the standardized matrix, **Quadrant I (High–High)** includes *Authenticity* (NFR2), *Cost* (NFR8), *Privacy* (NFR14), and *Protection* (NFR15), i.e., items that score above the median on both the rating and choice measures and thus form the verifiers' core consensus set.

Divergences appear in the off-diagonal quadrants. **Quadrant IV (High FI, Low BWS)** comprises *Verifiability* (NFR24), *Standard* (NFR20), and *Security* (NFR18), which receive relatively higher FI ratings than their revealed importance under forced-choice trade-offs. Conversely, **Quadrant II (Low FI, High BWS)** elevates *Consent* (NFR6), *Decentralization* (NFR9), and *Interoperability* (NFR11), indicating attributes that gain priority when respondents must make trade-offs. Several items are near the median cut lines, NFR9 and NFR1 near the II–III boundary, and NFR24 near all three quadrant boundaries, so small shifts could change their quadrant classification.

4.2.3.5 BWS Results – Issuers

A subset of issuer participants ($N = 37$) employed a Best–Worst Scaling (BWS) discrete choice experiment as a methodological complement to the direct rating approach (SQRI). The experiment produced repeated pairwise comparisons for 12 NFRs, with each item presented 105 times.

The model converged successfully, producing a full set of beta estimates with standard errors ranging from 0.19 to 0.27. Beta coefficients ranged from $\beta = 0.424$ (*Authenticity*, NFR2, rank 1) to $\beta = -1.380$ (*Interoperability*, NFR11, rank 12), with a M of $\beta = -0.491$ and a median of $\beta = -0.487$. Nine of the 12 items (75.0%) achieved statistical significance at $p < .05$: one item with a positive β value (indicating above-average importance) and eight items with negative β values (indicating below-average importance). Beta coefficients were transformed into importance shares (percentage of total utility), which sum to 100% across all items and facilitate intuitive interpretation of relative priorities. Importance shares ranged from 18.83% (*Authenticity*) to 3.10% (*Interoperability*), with the top five items accounting for 57.3% of total importance and the top ten items accounting for 92.9%.

Table 4.44: Complete Best-Worst Scaling results for all 12 NFRs (Issuers, $N = 37$).

Rank	NFR	β	SE	p	Share (%)	Best	Worst	Net	Norm
1	NFR2 – Authenticity	0.424	0.214	$< .05$	18.83	68	3	+65	0.62
2	NFR9 – Decentralization	0.000	–	–	12.32	8	67	-59	-0.56
3	NFR24 – Verifiability	-0.206	0.191	$> .05$	10.03	43	5	+38	0.36
4	NFR6 – Consent	-0.381	0.211	$> .05$	8.42	20	36	-16	-0.15
5	NFR5 – Compatibility	-0.467	0.229	$< .05$	7.73	13	52	-39	-0.37
6	NFR18 – Security	-0.475	0.239	$< .05$	7.67	45	7	+38	0.36
7	NFR8 – Cost	-0.499	0.209	$< .05$	7.48	2	54	-52	-0.50
8	NFR14 – Privacy	-0.552	0.238	$< .05$	7.09	32	23	+9	0.09
9	NFR21 – Transparency	-0.553	0.223	$< .05$	7.09	29	19	+10	0.10
10	NFR20 – Standard	-0.681	0.227	$< .01$	6.24	21	28	-7	-0.07
11	NFR15 – Protection	-1.121	0.252	$< .001$	4.02	22	8	+14	0.13
12	NFR11 – Interoperability	-1.380	0.274	$< .001$	3.10	12	13	-1	-0.01

Notes. Model estimates: β = utility coefficient, SE = standard error, p = significance threshold ($< .001$, $< .01$, $< .05$, or $> .05$), Share = importance share (%); Count metrics: Best/Worst = observed choice frequencies, Net = Best − Worst, Norm = standardized count score (Net/Exposure). All items had equal exposure of 105 presentations across respondents.

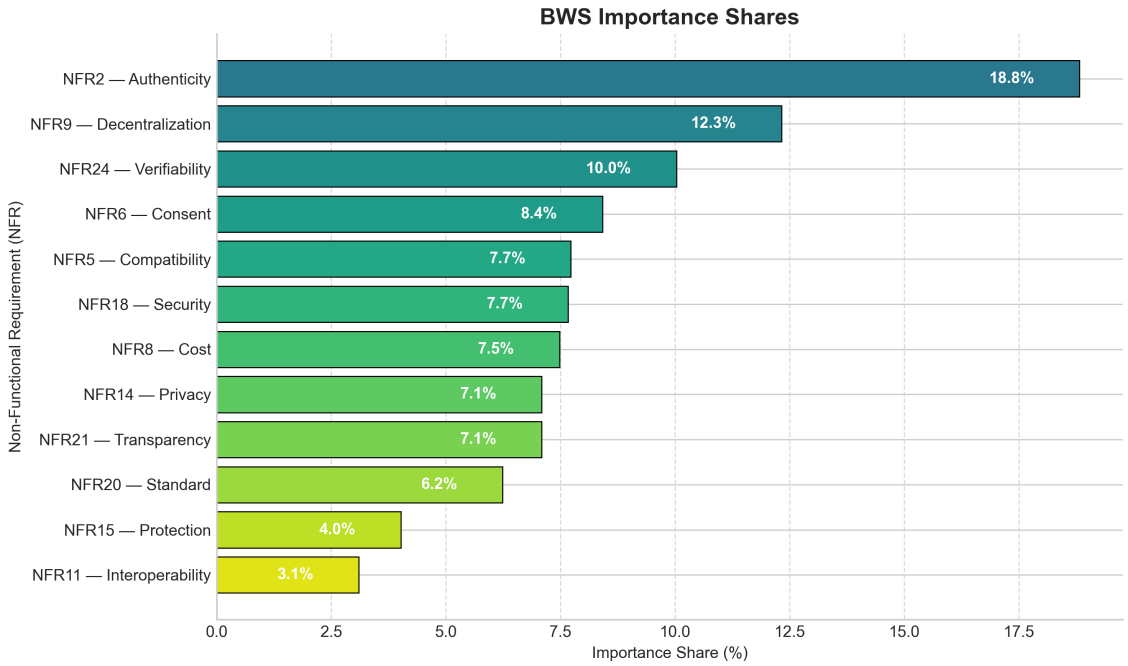

Figure 4.23: BWS importance shares ranked by model estimates. Importance shares sum to 100% across all items. Colors gradient from high (dark teal) to low (yellow) priority.

Table 4.44 and Figure 4.23 presents the complete BWS ranking of all 12 NFRs for issuers. *Authenticity* (NFR2) emerges as the highest priority by a substantial margin (rank 1, share = 18.83%, net = +65), followed by *Decentralization* (NFR9) (rank 2, share = 12.32%), *Verifiability* (NFR24) (rank 3, share = 10.03%), *Consent* (NFR6) (rank 4, share = 8.42%), and *Compatibility* (NFR5) (rank 5, share = 7.73%). These top five NFRs account for 57.3% of total importance. Notably, *Authenticity* is the only item with a positive β coefficient, indicating it was chosen as "best" substantially more often than "worst." The bottom tier comprises *Interoperability* (NFR11) (rank 12, $\beta = -1.380$, share = 3.10%), *Protection* (NFR15) (rank 11), and *Standard* (NFR20) (rank 10), with importance shares below 6.5%.

4.2.3.6 Alignment between SQRI and BWS – Issuers

To examine the alignment between rating-based and choice-based prioritization among issuers, standardized SQRI FI values were plotted against BWS importance scores using a common z-score scale. As above, axes are split by within-role medians (solid) with dashed means for reference, and medians are preferred due to observed ceiling/skew in Likert FI distributions. Figure 4.24 shows the resulting quadrant structure with NFR codes only and the same color scheme (QI blue, QII orange, QIII brown, QIV green).

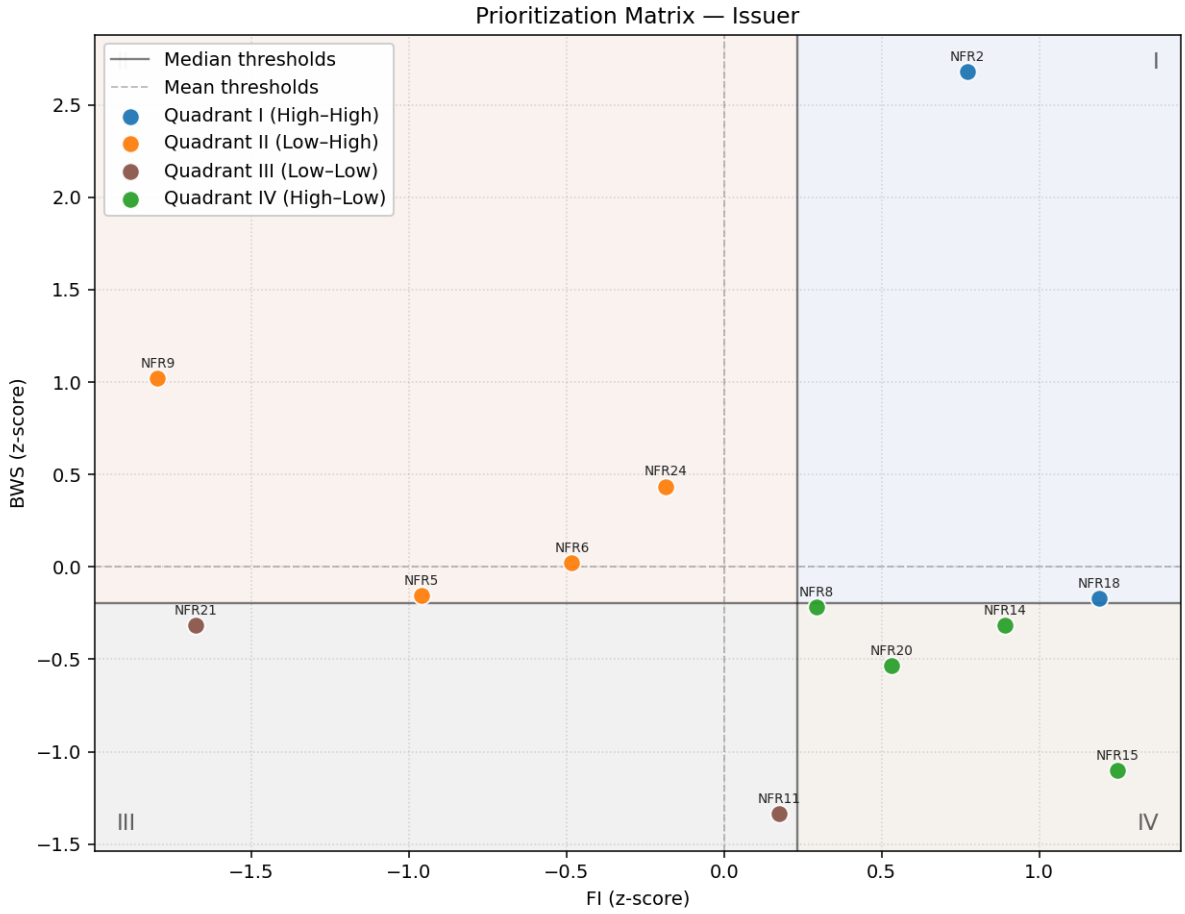

Figure 4.24: FI and BWS are z-scores; solid lines indicate median thresholds (classification), dashed lines indicate means (reference). Colors denote quadrants: I (High–High), II (Low–High), III (Low–Low), IV (High–Low).

The two approaches show weak, non-significant negative correlation (Pearson's $r = -0.135$, $p = 0.677$; Spearman's $\rho = -0.329$, $p = 0.297$), suggesting little convergence and pronounced method dependence of priorities. In the standardized matrix, **Quadrant I (High–High)** contains *Authenticity* (NFR2) and *Security* (NFR18), which remain above the median on both measures and thus define the issuers' core consensus.

Divergence is evident in the off-diagonals. **Quadrant IV (High FI, Low BWS)** includes *Cost* (NFR8), *Privacy* (NFR14), *Standard* (NFR20), and *Protection* (NFR15), indicating items with elevated ratings that drop in relative weight under forced-choice trade-offs. Conversely, **Quadrant II (Low FI, High BWS)** raises *Verifiability* (NFR24), *Consent* (NFR6), *Compatibility* (NFR5), and *Decentralization* (NFR9). Items close to the median cut lines include NFR5 and NFR21 (II–III boundary), NFR11 (III–IV boundary), and NFR8 and NFR18 (I–IV boundary), with NFR8 lying near the crosshair intersection; these should be interpreted cautiously, using the dashed mean lines as a sensitivity reference. Overall, the standardized matrix clarifies the small High–High consensus (*Authenticity*, *Security*) while highlighting the broader set of attributes whose priorities diverge between rating and trade-off tasks.

4.2.4 Cross-Role Analysis

The comparative analysis integrated the summary scores from the FI, PI_{rev}, and BWS datasets across stakeholder roles. As BWS and FI both quantified perceived importance, their measures allowed for direct comparison. In this analysis, FI and BWS scores were standardized before comparison to ensure consistency. PI focused on risk or loss when a quality was missing, and alignment was harder. The pipeline matched items to the same NFR codes, sorted them by scores for each method and role, and used Spearman rank correlations to show how similar the rankings were across roles and between FI and BWS within each role. The results were shown in easy-to-read Excel tables and charts (see Appendix B), including bar charts for the Top-10 BWS items and FI vs. BWS scatter plots. These outputs helped compare what mattered most to each group and clearly showed where priorities agreed or differed. PI results were still used to highlight which missing qualities were most concerning, but the main cross-method comparisons used FI and BWS for clarity and reliability.

The script implemented non-parametric tests to identify cross-role differences in FI and PI_{rev}, utilizing the Kruskal–Wallis test and calculating ε^2 effect sizes. Following significant omnibus results, pairwise Dunn tests with Bonferroni-adjusted p-values were conducted, reporting role-wise means and medians. For BWS, cross-role tests were conducted based on best/worst counts, reporting an omnibus chi-square across roles along with Holm-adjusted pairwise proportion z-tests, complete with directional indicators.

The comparative analysis of roles then explored heterogeneity among identity holders (users), issuers, and verifiers, statistically tested cross-role differences, and visualized areas where priorities aligned or diverged. Collectively, these outputs included concise statistical tables and publication-ready figures that addressed the core research questions and supported well-founded conclusions.

4.2.4.1 Cross-Role Comparison

To assess whether quality requirements prioritization patterns vary systematically across different stakeholder groups within the SSI ecosystem, this section presents a comparative analysis of NFRs rankings among three distinct user populations: identity holders (Users, $N = 86$), credential issuers (Issuers, $N = 37$), and credential verifiers (Verifiers, $N = 27$). The analysis examines 13 overlapping NFRs common across all roles using three complementary measurement approaches: SQRI Likert ratings (FI, PI) and BWS choice frequencies.

Three stakeholder samples completed the SQRI questionnaire assessing 13 overlapping NFRs that are common across all roles: *Accessibility* (NFR1), *Authenticity* (NFR2), *Compatibility* (NFR5), *Consent* (NFR6), *Cost* (NFR8), *Decentralization* (NFR9), *Interoperability* (NFR11), *Privacy* (NFR14), *Protection* (NFR15), *Security* (NFR18), *Standard* (NFR20), *Transparency* (NFR21), and *Verifiability* (NFR24). Users additionally evaluated 11 role-specific NFRs (24 total), while Issuers and Verifiers assessed only the 13 common items. Missing data were minimal across all samples (less than 2%), with listwise deletion applied per item. All analyses utilize FI and PI_{rev} Likert scales (1–5), where

higher scores indicate greater importance or problem severity. A subset of each role also completed Best-Worst Scaling (BWS) tasks: Identity Holders (Users) ($N = 78$), Issuers ($N = 35$), and Verifiers ($N = 24$).

Figures 4.25 and 4.26 present three-panel prioritization matrices showing the relationship between FI and PI_{rev} and between FI and BWS importance shares, respectively, across all three stakeholder roles. Each panel displays data for one role on unified scales, enabling direct visual comparison of prioritization patterns. Points in the upper-right quadrant (high FI, high PI_{rev} or BWS) represent consensus high-priority items, while lower-left quadrants indicate lower-priority requirements.

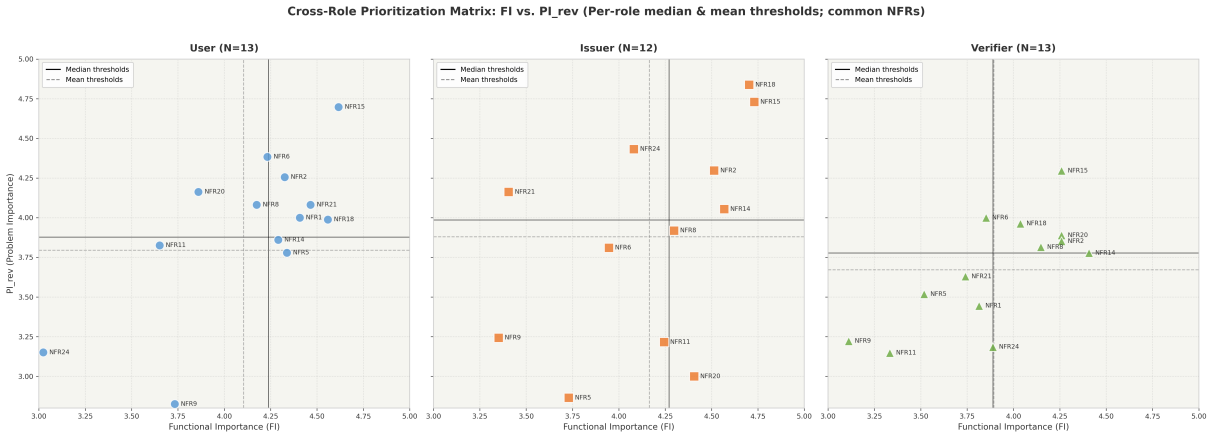

Figure 4.25: Three panels show Identity Holders (Users) (N=13 common NFRs), Issuers (N=12), and Verifiers (N=13) on unified scales.

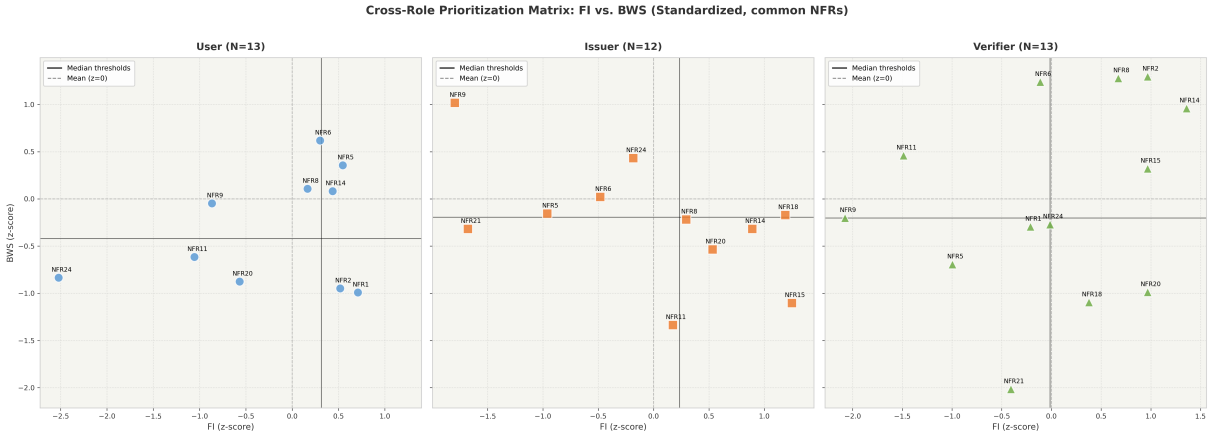

Figure 4.26: Three panels show Identity Holders (Users) (N=13), Issuers (N=12), and Verifiers (N=13) on unified scales.

Effect Size Summary
Across all Kruskal-Wallis tests, effect sizes varied substantially. For FI ratings, ε^2 ranged from 0.003 to 0.142, with two items (*Transparency* and *Verifiability*) achieving the large effect threshold ($\varepsilon^2 \geq 0.14$). For PI_{rev} ratings, ε^2 ranged from 0.001 to 0.207, with *Verifiability* (NFR24, $\varepsilon^2 = 0.207$) exceeding the large effect threshold. Rank-biserial effect sizes in pairwise comparisons ranged from $|r| = 0.006$ to $|r| = 0.611$, with 10 of 22 pairwise comparison.

Consensus Items
Five NFRs showed no significant cross-role differences in either FI or PI_{rev} ratings after Holm correction: *Privacy* (NFR14), *Protection* (NFR15), *Consent* (NFR6), *Authenticity* (NFR2), and *Cost* (NFR8). These items represent areas of stakeholder consensus, with all three roles rating them similarly in both FI and PI_{rev} ($p_{\text{adj.}} = 1.00$ for all comparisons).

Within-Role FI-PI_{rev} Correlations
Within-role correlations between FI and PI_{rev} rankings were computed to assess alignment between FI and problem severity perceptions within each stakeholder group (Table 4.45).

Table 4.45: Within-role Spearman correlations between FI and PI_{rev} rankings.

Role	N **(NFRs)**	**Spearman** ρ	p**-value**
Users	24	0.561	.004
Issuers	12	0.539	.071
Verifiers	13	0.641	.018

Notes. Correlations between FI and PI_{rev} mean rankings within each role. Users and Verifiers show significant positive correlations at $\alpha = .05$.

Kruskal-Wallis Tests: Functional Importance
Kruskal-Wallis rank-sum tests were conducted for each of the 13 common NFRs to test whether FI ratings differ significantly across the three stakeholders formally. This non-parametric approach is suitable for ordinal Likert data with unequal group sizes and does not require normality assumptions. The test statistic H was computed for each item, with Holm correction applied across all 13 tests to control family-wise error rate at $\alpha = .05$. Effect sizes were quantified using ε^2, a rank-based measure of association strength.

Table 4.46: Kruskal-Wallis tests for role differences in FI ratings (13 common NFRs), ranked by effect size ε^2.

NFR	H	p **(raw)**	$\boldsymbol{\varepsilon^2}$	$\boldsymbol{p}_{\text{adj.}}$
Transparency (NFR21)	22.91	$< .001$	0.142	$< .001$
Verifiability (NFR24)	22.82	$< .001$	0.142	$< .001$
Compatibility (NFR5)	19.35	$< .001$	0.118	.001
Accessibility (NFR1)	9.49	.002	0.076	.043
Interoperability (NFR11)	10.03	.007	0.055	.133
Security (NFR18)	9.46	.009	0.051	.168
Decentralization (NFR9)	6.66	.036	0.032	.643
Standard (NFR20)	6.10	.047	0.028	.806
Protection (NFR15)	4.04	.132	0.014	1.00
Privacy (NFR14)	3.09	.213	0.007	1.00
Consent (NFR6)	2.49	.288	0.003	1.00
Authenticity (NFR2)	1.20	.548	–	1.00
Cost (NFR8)	0.37	.832	–	1.00

Notes. Three-group comparison (Identity Holder (User), Issuer, Verifier) for FI block. Holm correction applied across 13 tests. ε^2 omitted for non-significant items with negative preliminary estimates. *Accessibility* (NFR1) tested with only two groups (Identity Holders (Users), Verifiers) as Issuers did not assess this item.

Kruskal-Wallis Tests: Problem Importance

Kruskal-Wallis rank-sum tests were also conducted for PI_{rev} ratings across roles, with Holm correction applied across all 13 tests. Effect sizes were quantified using ε^2.

Table 4.47: Kruskal-Wallis tests for role differences in PI_{rev} ratings (13 common NFRs), ranked by effect size ε^2.

NFR	H	p **(raw)**	$\boldsymbol{\varepsilon^2}$	$\boldsymbol{p}_{\text{adj.}}$
Verifiability (NFR24)	32.40	$< .001$	0.207	$< .001$
Standard (NFR20)	20.60	$< .001$	0.126	$< .001$
Security (NFR18)	19.19	$< .001$	0.117	.001
Compatibility (NFR5)	16.25	$< .001$	0.097	.006
Accessibility (NFR1)	9.10	.003	0.073	.051
Interoperability (NFR11)	10.30	.006	0.056	.110
Protection (NFR15)	7.46	.024	0.037	.432
Consent (NFR6)	6.54	.038	0.031	.646
Transparency (NFR21)	4.48	.107	0.017	1.00
Decentralization (NFR9)	4.01	.135	0.014	1.00
Authenticity (NFR2)	2.33	.312	0.002	1.00
Cost (NFR8)	2.07	.354	0.001	1.00
Privacy (NFR14)	1.57	.457	–	1.00

Notes. Three-group comparison for PI_{rev} block. Holm correction applied across 13 tests. ε^2 omitted for non-significant items with negative preliminary estimates.

Post-Hoc Pairwise Comparisons

Post-hoc pairwise Dunn tests with Bonferroni-adjusted p-values were conducted for items with significant omnibus results.

Table 4.48: Post-hoc pairwise Dunn-Bonferroni tests for significant FI items.

NFR	Comparison	z	r	p_{Bonf}		Direction
Accessibility (NFR1)	User vs. Verifier	3.08	0.290	.002	**	User > Verifier
Transparency (NFR21)	Issuer vs. User	−4.36	0.356	<.001	***	User > Issuer
	User vs. Verifier	3.07	0.251	.004	**	User > Verifier
	Issuer vs. Verifier	−0.71	0.058	.477		
Verifiability (NFR24)	Issuer vs. User	4.31	0.352	<.001	***	Issuer > User
	User vs. Verifier	−3.14	0.256	.003	**	Verifier > User
	Issuer vs. Verifier	0.61	0.050	.542		
Compatibility (NFR5)	Issuer vs. User	−3.51	0.287	.001	**	User > Issuer
	User vs. Verifier	3.50	0.285	.001	**	User > Verifier
	Issuer vs. Verifier	0.32	0.026	.748		

Notes. z = standardized test statistic from post-hoc Dunn-Bonferroni tests; r = rank-biserial effect size (ranges from 0 to 1). ***$p < .001$, **$p < .01$, *$p < .05$ (Bonferroni correction applied within each NFR's pairwise family). Direction indicates which group has significantly higher ratings; empty cells indicate no significant difference. Accessibility tested only for User vs. Verifier (Issuers did not assess NFR1).

Table 4.49: Post-hoc pairwise Dunn-Bonferroni tests for significant PI_{rev} items.

NFR	Comparison	z	r	p_{Bonf}		Direction
Verifiability (NFR24)	Issuer vs. User	5.52	0.451	<.001	***	Issuer > User
	Issuer vs. Verifier	4.16	0.340	<.001	***	Issuer > Verifier
	User vs. Verifier	−0.14	0.012	.887		
Standard (NFR20)	Issuer vs. User	−4.54	0.370	<.001	***	User > Issuer
	Issuer vs. Verifier	−2.35	0.192	.038	*	Verifier > Issuer
	User vs. Verifier	1.35	0.110	.177		
Security (NFR18)	Issuer vs. User	4.20	0.343	<.001	***	Issuer > User
	Issuer vs. Verifier	3.32	0.271	.002	**	Issuer > Verifier
	User vs. Verifier	0.07	0.006	.945		
Compatibility (NFR5)	Issuer vs. User	−4.03	0.329	<.001	***	User > Issuer
	Issuer vs. Verifier	−2.27	0.185	.046	*	Verifier > Issuer
	User vs. Verifier	0.99	0.081	.323		

Notes. z = standardized test statistic from post-hoc Dunn-Bonferroni tests; r = rank-biserial effect size. ***$p < .001$, **$p < .01$, *$p < .05$. Direction indicates which group rates the problem as more severe; empty cells indicate no significant difference.

Cross-Role Ranking Correlations

Spearman rank correlations were computed on mean FI, PI_{rev}, and BWS importance share rankings for the 13 common NFRs to assess overall alignment in NFR prioritization between roles.

Table 4.50: Spearman rank correlations between role-specific NFR rankings (13 common items).

Metric	User vs. Issuer	User vs. Verifier	Issuer vs. Verifier
FI mean	0.406 ($p = .191$)	0.227 ($p = .457$)	0.817 ($p = .001$)
PI_{rev} mean	0.273 ($p = .390$)	0.820 ($p < .001$)	0.371 ($p = .236$)
BWS share	-0.259 ($p = .417$)	-0.225 ($p = .459$)	0.245 ($p = .443$)

Notes. Correlations computed on ranks (1–13) for each metric within each role. Bold indicates $p < .05$. $N = 13$ NFRs for all comparisons.

Within-Role Correlation Analysis

To assess the convergence and divergence across prioritization methods within each stakeholder group, both Pearson and Spearman correlations were computed between the three measurement approaches: FI, PI_{rev}, and BWS importance shares. Pearson correlations assess linear relationships between variables, while Spearman rank correlations examine monotonic relationships and are more robust to outliers and non-normal distributions. Table 4.51 presents these within-role correlations, revealing substantial variation in method alignment across stakeholder groups.

Table 4.51: Within-role correlations between prioritization methods.

Role	Comparison	N	Pearson r	p	Spearman ρ	p	Sig.
Users	FI vs. PI_{rev}	24	0.709	<.001	0.561	.004	*** / **
Users	FI vs. BWS	24	0.458	.025	0.373	.073	* / ns
Issuers	FI vs. PI_{rev}	12	0.481	.113	0.539	.071	ns / ns
Issuers	FI vs. BWS	12	-0.135	.677	-0.329	.297	ns / ns
Verifiers	FI vs. PI_{rev}	13	0.738	.004	0.641	.018	** / *
Verifiers	FI vs. BWS	13	0.272	.369	0.304	.313	ns / ns

Notes. Pearson correlations assess linear relationships; Spearman correlations assess monotonic rank-order relationships. Sig. column shows Pearson / Spearman significance. FI = Functional Importance; PI_{rev} = Problem Importance (reversed); BWS = Best-Worst Scaling. ns = not significant, * $p < .05$, ** $p < .01$, *** $p < .001$.

Chapter 5

Discussion

In this chapter, the results will be discussed. This study sought to empirically identify which quality requirements are prioritized by different stakeholder groups in SSI ecosystems and to assess whether participants clearly understood the mapping between functional descriptions and quality requirements. Using a mixed-methods approach combining SQRI Likert-scale ratings, BWS, and visual prioritization matrices, the analysis addresses two core research questions: (i) which qualities are important for each category of users, and (ii) are the described functionalities and their mapping to quality requirements precise to users.

This study provides empirical evidence on the quality requirements and priorities for SSI systems across three distinct stakeholder groups: identity holders (users), credential issuers, and credential verifiers. The statistical analysis revealed both areas of consensus and significant divergence in how different roles prioritize NFRs.

5.1 Stakeholder Priorities for NFRs

5.1.1 Identity Holders (Users)

The questionnaire results reveal clear patterns in how different stakeholders prioritize quality requirements. The identity holders (users) generally rated most NFRs as having neutral importance. Users consistently rated *Protection* (NFR15), *Security* (NFR18), *Recoverability* (NFR16), *Control* (NFR7), and *Transparency* (NFR21) as the most functionally important requirements, with mean FI scores ranging from 4.47 to 4.62. This pattern reflects end-users' fundamental concerns about data safety and personal agency, and maintaining agency over data disclosure. On the PI_{rev} scale, *Protection* (NFR15), *Consent* (NFR6), *Authenticity* (NFR2), *Control* (NFR7), and *Standard* (NFR20) emerged as the top priorities. This ranking is reflected in the Friedman mean-rank table for users (Table 4.6).

The Friedman tests indicate that users clearly differentiate between NFRs in both frames (FI and PI_{rev}). Agreement among respondents is moderate and is stronger when the scenario highlights problems or failures, suggesting that such contexts make user preferences more distinct. In both blocks, *Protection* consistently ranks highest. In FI, *Security*, *Control*, *Transparency*, and *Recoverability* follow, while in PI_{rev}, *Consent* and *Authenticity* become more prominent. These results align with expert benchmarks, which typically identify *Security* or *Protection* and *Verifiability* or *Authenticity* as top priorities, but they also reveal that end-users value *Control* and *Transparency* in everyday settings [98]. This finding supports the idea that user trust is shaped by evident protections and transparent data management, not solely by technical or cryptographic assurances [88]. Distinct pairwise effects, such as the high priority of *Protection* over *Representation* or *Verifiability*, demonstrate that these rank differences are meaningful and not due to random variation. The observed decline in *Usability* and *Accessibility* in PI_{rev} suggests that when potential risks are emphasized, users tend to value strong safeguards above convenience, which is a typical pattern in digital identity systems where *Privacy* and *Security* dominate user preferences [106].

Identity holders demonstrated a strong consensus on core security and usability-related qualities. *Protection* of identity data emerged as the most critical requirement, with the highest FI mean value of 4.62. Over half of the users gave *Protection* the highest possible importance, reflecting an unequivocal demand that their personal data be safeguarded from breaches and misuse. *Security* (the technical safety of the system) was nearly as important (mean 4.56), reinforcing the view that users consider secure credentials and infrastructure a fundamental need. Notably, users also highly prioritized the ability to recover their identity (*Recoverability* mean 4.56), a recognition that losing access to keys or accounts is catastrophic in SSI, making backup and restoration mechanisms essential [106]. In addition to these security-oriented qualities, users placed great value on having control over who accesses their data (*Control* mean 4.51) and on transparency in the system's operations (*Transparency* mean 4.47). These top priorities indicate that SSI users seek both the autonomy to manage their identity and the assurance that the system's handling of their data is open and auditable. Furthermore, users want an SSI wallet that keeps their data safe, preserves privacy, and gives them agency [86]. For example, over half of respondents assigned *Protection* the highest importance rating, indicating an unequivocal demand for robust safeguards against data breaches. High rankings for accessibility (mean 4.41) and usability (mean 4.36) further underscore the importance of an SSI wallet that is available whenever needed and easy to use in everyday tasks.

The user community expects an SSI solution that is secure by design, preserves privacy, and remains user-centric, aligning closely with the envisioned SSI principles of user control and data protection. At the same time, qualities like decentralization and system independence were relatively less important to users: for example, *Decentralization* (eliminating any central authority) received a middling importance (mean 3.73, near neutral). For example, *Autonomy* (having no external authority control) and *Decentralization* were among the lowest-ranked by users (FI ranks 17 and 18 of 24), which suggests that while *Autonomy/Decentralization* are core principles of SSI, end-users do not perceive them as immediately critical compared to security and privacy. Even though the quality *Autonomy* was assigned as a primary responsibility, it ranked 17th compared to NFR *Security*, which ranked 2nd in FI. Notably, when considering problem scenarios (PI ratings), users

showed heightened concern for specific issues that were not top-of-mind in abstract importance. *Protection* remained the most alarming scenario for users (PI_{rev} mean 4.70, rank 1). However, *Consent* violations emerged as a close second (PI_{rev} mean 4.38, rank 2) despite *Consent* being mid-ranked on FI, indicating that users may take consent for granted in everyday use. However, a scenario in which personal data is used without permission is highly distressing. Similarly, *Authenticity* (the risk of falsified identity data) was rated more concerning in the problem context (PI_{rev} rank 3) than its FI rank (8), suggesting users recognize the importance of authentic credentials when faced with the prospect of fraud. Overall, the user perspective prioritized qualities that affect their personal security, privacy, and ability to control and recover their data, with BWS confirming that *Security*, *Protection*, *Transparency*, and *Control* form a consensus top tier of user priorities. In the BWS section (which forces trade-offs), these four qualities accounted for a large share of importance, and *Consent* also entered the users' top five when respondents had to choose trade-offs. The alignment of BWS with rating-based results for users was strong for the aforementioned top-four NFRs, reinforcing that users consistently view data safety and agency as paramount.

These priorities align with identity holders' assigned primary responsibilities [98]. Users hold primary responsibility for *Protection*, *Control*, *Recoverability*, *Transparency*, and *Accessibility*, which explains their high ratings. However, the low ranking of *Autonomy* and *Representation*, despite being assigned as primary responsibilities, reveals a notable gap: users may not recognize or value these foundational SSI principles, potentially because they perceive them as abstract system properties rather than actionable user duties. Many of these top-rated NFRs are ones inherently owned by users, making them direct beneficiaries of consent, control, and data protection. When a primary quality is perceived as indirect or abstract, users may not emphasize it unless its absence creates a clear problem.

5.1.2 Credential Verifiers

The survey of verifiers revealed a distinct priority profile reflecting their specific operational needs and constraints. In verifiers' FI ratings, *Privacy* of user data was surprisingly the highest-rated quality (FI mean 4.41, rank 1), indicating that verifiers (who consume identity data) still highly value privacy protections, likely to ensure user trust and compliance. Close behind were *Authenticity* of credentials and *Protection*, tied with *Standard*, all with a mean FI $\approx$ 4.26 (ranks 2–4). This suggests that verifiers strongly prioritize that identity data is trustworthy and secure, and that systems follow standards (interoperability standards make verification easier). Cost-efficiency (*Cost*, FI 4.15, rank 5) and *Security* (4.04, rank 6) also ranked high for verifiers, consistent with businesses caring about the performance and expenses of verification processes. Notably lower in their FI were end-user-centric qualities, e.g., *Accessibility* and *Transparency* ranked in the bottom half (FI ranks 9–10), reflecting that verifiers see these as tertiary and secondary responsibilities. When considering PI, verifiers' top concerns shifted: a breach of *Protection* was rated the most critical scenario (PI_{rev} mean 4.30, rank 1), implying that verifiers are acutely alarmed by any misuse or compromise of identity data under their watch. *Consent* jumped from 8th in FI to the second-highest concern in PI_{rev} (mean 4.00). Verifiers may

not emphasize obtaining user consent day to day, but the idea of using data without consent is highly concerning, likely due to legal and ethical implications. *Security* issues (e.g., system outages or breaches) ranked third among concerns (PI_{rev} mean 3.96). Meanwhile, *Standard* and *Authenticity* remained important in both frames (each with PI_{rev} means of ∼3.85–3.89 and top-5 ranks). Verifiers considered certain qualities to be much less urgent in specific scenarios; for instance, *Verifiability* (the ease of verifying credentials) dropped to near the bottom of the concern list (PI_{rev} rank 12) despite a moderate FI rank, indicating that verifiers might assume verification processes will work and thus did not fear their absence.

Verifiers' FI ranks are concentrated on *Privacy*, *Protection*, *Standard*, *Authenticity*, and *Cost*, with *Security* ranking just below these. In the PI_{rev} frame, *Protection*, *Consent*, and *Security* become the top concerns, while *Verifiability* is ranked lower and overall agreement among verifiers decreases. This result is consistent with the expected verifier profile, as these actors are responsible for processing personal data lawfully, preventing breaches, and maintaining efficient, compliant verification workflows. Previous research has found that organizations prioritize standards and verification properties because broad ecosystem acceptance and auditability support operational effectiveness [64, 104]. The small number of significant PI_{rev} pairs and lower Kendall's W both indicate that verifiers operate in more diverse contexts, which matches software engineering findings that stakeholder groups vary in how they prioritize NFRs across scenarios, and that results are influenced by the choice of prioritization method [57, 75].

Verifiers hold primary responsibility for *Consent*, *Privacy*, and *Verifiability* [98]. Empirical priorities partially align with these assignments, as *Privacy* (rank 1) and *Authenticity* (rank 2) ranked highly. However, *Verifiability* was assigned as primary yet ranked low in PI (PI rank 12), suggesting verifiers may perceive verification as an automatic system capability rather than an active responsibility. Conversely, *Security* was assigned a tertiary responsibility but ranked highly (FI rank 6), indicating a misalignment between role definitions and practical concerns.

In the BWS trade-off results for verifiers, practical considerations rose to the top: the highest-priority forced-choice option was *Authenticity*, followed by *Cost*, *Consent*, *Privacy*, and *Interoperability*. This reveals that when push comes to shove, verifiers tend to favor requirements that ensure credentials are genuine, exchanges are economical, and privacy is protected. Interestingly, *Security* and *Standard* fell to the bottom of the verifiers' BWS ranking (ranks 12–13), even though they had endorsed those in abstract ratings. This divergence suggests that verifiers conceptually acknowledge the importance of security and standards. However, when forced to prioritize, they place relatively greater weight on immediate operational factors (such as *Authenticity* and *Cost*) than on broader principles, like standard compliance. It underscores a potential gap between what verifiers say is important in theory and the trade-offs they would actually make.

5.1.3 Credential Issuers

Issuers showed the most concentration on data integrity and security-related qualities. Issuers gave exceptionally high FI ratings to *Protection* (mean 4.73, rank 1) and *Security* (4.70, rank 2), with virtually all issuers rating these as extremely important. These two qualities form the bedrock for issuers: protecting issued identity data and ensuring secure handling are seen as essential. *Privacy* was also very highly rated by issuers (FI 4.57, rank 3), reflecting issuers' responsibility to safeguard personal data they issue. *Authenticity* (ensuring credentials' truthfulness) and *Standard* compliance (following standards/protocols) were ranked 4 and 5, respectively (FI 4.51 and 4.41). Issuers prioritize making credentials broadly trustworthy and usable across systems, not surprisingly, since an issuer's reputation relies on credentials being accepted and lasting. In contrast, issuers placed lower importance on qualities such as *Interoperability* (FI 4.24, mid-tier) and *Transparency* (FI 3.41, rank 11). *Transparency* was the lowest FI for issuers, implying they do not feel a pressing need for system openness from their perspective (perhaps assuming internal processes are sufficient).

However, when evaluating PI items, issuers' priorities shifted in revealing ways. *Security* breaches were the single most frightening scenario for issuers (PI mean 4.84, rank 1), reaffirming that nothing worries credential issuers more than a security failure that could compromise their issued credentials or systems. *Protection* failures (e.g., data misuse) were nearly as concerning (PI 4.73, rank 2). Interestingly, *Verifiability*, the ability of third parties to verify credentials, jumped from a low FI (rank 8) to the third-highest PI concern (mean 4.43). This suggests issuers might take verifiability for granted in their design, but if credentials cannot be verified, it becomes an alarming scenario that threatens the utility of their role. *Transparency* showed the opposite pattern: issuers initially gave it little weight, but a lack of transparency scenario was moderately concerning (PI 4.16, rank 5), which suggests that while issuers do not emphasize building transparency, they do acknowledge that an opaque system can lead to trust issues or user backlash in worst-case situations. Meanwhile, *Standard* plummeted in issuers' PI rankings (from FI rank 5 to PI rank 11). Apparently, an absence of common standards, while suboptimal, is not seen as an immediate "concern" scenario for issuers, perhaps because they focus on their own standards compliance and view a lack of global standards as a tolerable, if unfortunate, state.

Issuers display clear distinctions in their NFR rankings, particularly in the $\mathrm{PI}_{\mathrm{rev}}$ frame, where agreement is higher, *Security* and *Protection* are prioritized, and *Verifiability* receives much greater emphasis. This pattern reflects issuers' responsibility for ensuring credential integrity and supporting revocation, as well as the practical need for credentials to remain verifiable even if the issuer is unavailable. Academic literature highlights *Authenticity* and *Verifiability* as essential prerequisites for a reliable digital identity ecosystem, which accounts for the prominent contrasts issuers make between highly ranked and lower-priority items [64, 98]. The observed drop in *Standards* under $\mathrm{PI}_{\mathrm{rev}}$ suggests that while non-compliance is considered costly, it is not viewed as severely as a security breach, a trend documented in frameworks that distinguish between regulatory conformance and safety-critical risks [104].

Issuers are primarily responsible for *Authenticity*, *Compatibility*, *Interoperability*, and

Standard, according to the NFR mapping conducted earlier in this study. Empirical priorities strongly align with *Authenticity* (rank 4) and *Standard* (rank 5), confirming issuers recognize their role as guarantors of credential integrity [59, 98]. However, *Interoperability* and *Compatibility* received lower ratings, suggesting that issuers prioritize internal credential quality over cross-system compatibility. Conversely, *Security* was assigned a tertiary responsibility yet ranked second (FI 4.70), indicating that issuers perceive *Security* as a core operational concern despite its classification also as a system-level property. This gap indicates issuers may conflate their direct security duties with broader system security, potentially overlooking that their primary responsibility centers on authentic credential creation rather than comprehensive system protection.

Issuers' empirical priorities center on ensuring the credentials they issue are secure, protected, and can be trusted and verified by others. Secondary considerations (like user-facing transparency or cross-compatibility) receive attention mainly when considering the implications of their failure. This aligns with issuers' primary role as guarantors of data integrity. Indeed, the BWS results for issuers (though covering only 12 key NFRs) confirm a security-centric profile at the top. For instance, *Security* and *Protection* were among the highest-share attributes in issuers' BWS ranking. Conversely, qualities like *Compatibility* (with legacy systems) ranked lowest among issuers in both ratings and BWS (FI rank 10; PI_{rev} rank 12), indicating that backward compatibility is not a priority for issuers relative to other concerns.

Overall, each stakeholder group's priorities align with their role's responsibilities and risks: users prioritize personal data protection and control, verifiers prioritize trustworthiness and verification efficiency, and issuers prioritize the security and reliability of the credentials they issue.

5.1.4 Visual Prioritization Patterns

The prioritization matrices (Figures 4.10, 4.2.2.7, 4.18) plot each NFR's mean FI against its mean PI (reversed) to create a 2×2 framework dividing requirements into four quadrants:

- **Quadrant I (High FI, High PI_{rev}):** All three roles unanimously positioned *Security*, *Protection*, and *Authenticity* here. Wilcoxon signed-rank tests confirmed all High-High quadrant NFRs were rated significantly above the neutral midpoint ($p < .05$). These findings align with established guidance that privacy, *Security*, and trust require early attention in system design [105]. [106] similarly identify *Security* as a foremost SSI quality requirement. *Privacy* is classified in Quadrant I for both issuers and verifiers, indicating high FI and High PI. For users, however, Privacy shifts into the High FI / Low PI quadrant under the median split and is positioned exactly on the border, suggesting a tendency toward higher prioritization but not clearly joining the core consensus group. *Cost* is classified in Quadrant I for verifiers, signifying High FI and High PI in this group. For issuers, it appears in the High FI / Low PI quadrant, while for users, it falls into the Low FI / High PI quadrant. This pattern matches the RE literature that security, privacy/trust, and proof of

authenticity anchor the quality baseline for socio-technical systems [105, 106], while the exact emphasis varies by role [36, 57, 75].

- **Quadrant II (Low FI, High PI_{rev}):** Flags latent risks where stakeholders undervalue specific attributes until their absence becomes problematic.
  - *Users: Standard*, *Consent*, *Cost*, and *Autonomy*, attributes reflecting latent risks that users primarily notice in the event of failures.
  - *Issuers: Verifiability* and *Transparency*, where issues become salient even if their FI is below the median.
  - *Verifiers: Consent*, which presents legal and ethical risks if violated, despite mid-tier FI ratings.

  Standard varies significantly across stakeholders: issuers position it in High FI-Low PI, verifiers in High FI-High PI, and users in Low FI-High PI. Verifiers experience standardization as both important and problematic, while users encounter problems without recognizing their underlying importance. *Transparency* shows pronounced differences across roles: users place it in the High FI–High PI quadrant, viewing it as both highly important and problematic; issuers position it in the Low FI–High PI quadrant, indicating it is problematic but undervalued in terms of FI; and verifiers assign it to the Low FI–Low PI quadrant, suggesting it is not considered a priority.

- **Quadrant III (Low FI, Low PI_{rev}):** Attributes rated low in both abstract importance and perceived problem salience are concentrated in this quadrant. Notably, *Interoperability* is located in Quadrant III for all three roles under the median split. In addition, users assign *Portability*, *User Experience*, *Availability*, *Decentralization*, *Verifiability*, *Representation*, and *Existence* to the same quadrant. Issuers assign *Interoperability*, *Compatibility*, *Consent*, and *Decentralization* to this quadrant. Verifiers place *Accessibility*, *Transparency*, *Compatibility*, *Interoperability*, and *Decentralization* here as well. This finding stands in contrast to segments of the SSI literature that position *Interoperability* and *Verifiability* as universally high priorities [106]. Our results indicate these attributes are not consistently rated as top-tier by all stakeholders, and often fall outside Quadrant I except where they are directly essential for specific role workflows, echoing technical integration observations reported by [104].

- **Quadrant IV (High FI, Low PI_{rev}):** This quadrant includes qualities regarded as important "in principle" but not currently viewed as problematic. For users, *Usability*, *Compatibility*, *Privacy*, and *Persistence* are located here; *Privacy* is positioned at the borderline to Quadrant I, indicating a tendency toward higher prioritization in both FI and PI. For issuers, *Standard* and *Cost* fall into this group. For verifiers, *Verifiability* is the only item found in Quadrant IV, lying at the boundary with Quadrant III. The previous mention of *Interoperability* near Quadrant IV is corrected: it is consistently ranked in Quadrant III. These role-specific allocations reflect prior findings that stakeholders prioritize the same NFRs differently depending on their operational responsibilities and constraints [57, 75, 98].

This layout follows standard priority matrices in requirements engineering, distinguishing always-critical NFRs from context-dependent ones. Results confirm that Quadrant I

qualities reflect genuine stakeholder consensus, while borderline items in lower quadrants lack strong agreement on importance.

The matrices reveal role-based differences, consistent with previous research showing that stakeholders prioritize quality attributes according to their responsibilities [36]. Each group emphasized qualities linked to their operational context. Under the median split, users placed *Protection*, *Security*, *Control*, *Transparency*, *Accessibility*, *Recoverability*, *Authenticity*, and *Single Source* in Quadrant I, reflecting user experience and trust factors that are significant in daily use and comparatively well delivered. In contrast, *Privacy* and *Persistence* for users fall into High FI and Low PI, while *Consent* and *Standard* are classified as Low FI and High PI. Issuers and verifiers ranked user-centric NFRs such as *Accessibility*, *Compatibility*, and *User Experience* lower, indicating these attributes are less pivotal to their tasks. Borderline and divergent cases highlight that priorities are context-dependent; mid-ranked and divergent qualities should not be overlooked simply because they appear less important for certain stakeholder groups.

To compare FI with BWS, standardized FI and BWS scores were analyzed, using a median split to assign items to quadrants (see Figures 4.20, 4.22, 4.24). Medians were chosen over means because FI items are measured on a 5-point Likert scale, which exhibited asymmetry and ceiling effects in the data. The median offers a more robust classification approach for skewed ordinal data. Using medians to divide the axes is appropriate when the distributions are skewed, ceiling-prone, and contain tied values. Median-based thresholds are robust against differences in scale use, as supported by [34, 94]. Likert-based FI ratings demonstrated a ceiling effect, with many respondents giving NFRs high scores, compressing differences at the top. [55] note that Likert scales often produce ceiling effects. BWS forced trade-offs and spread out priorities.

Correlation analysis across all three prioritization methods confirms these findings (Section 4.2.4.1). FI and PI showed strong positive correlations among users and verifiers, indicating consistent perceptions across abstract and experiential importance dimensions. Issuers, in contrast, showed no significant correlation. The FI-BWS correlations also varied: users demonstrated moderate agreement ($r = 0.46$, $p = .025$), while issuers ($r = -0.13$, $p = .677$) and verifiers ($r = 0.27$, $p = .369$) showed weak, non-significant relationships. These results highlight the need for multi-method triangulation and demonstrate that stakeholder groups have fundamentally different priority structures.

For identity holder (users), *Security*, *Protection*, *Transparency*, and *Control* occupy High FI and High BWS quadrants. [50] found that improving data control and transparency boosts customer satisfaction, which helps explain why *Transparency* is positioned in Quadrant I. In contrast, convenience items such as *Accessibility*, *Authenticity*, and *Recoverability* tend to fall into High FI and Low BWS when stakeholders face trade-offs. *Recoverability*, *Accessibility*, and *Authenticity* are positioned in High FI and Low BWS for users; when required to make trade-offs, users tend to prioritize front-line safeguards such as *Security*, *Protection*, *Transparency*, and *Control* over second-line mitigations like recovery. *Authenticity* may also be partially covered by *Security* and *Protection*, reducing the need for separate prioritization. Issuers prioritize *Authenticity* in both dimensions, with *Verifiability* and *Decentralization* found in Low FI and High BWS, highlighting latent priorities that FI scores do not fully capture.

For verifiers, *Authenticity*, *Privacy*, *Cost*, and *Protection* are the main High FI–High BWS items, while *Consent* lands in Low FI–High BWS, reflecting hidden risks even when functional ratings are moderate. These findings indicate that Quadrant I items consistently represent the most important requirements across measurement approaches. The BWS method further clarifies where Likert items are overstated and underscores hidden constraints that may surpass surface ratings. Employing both methods identifies which requirements genuinely matter when stakeholders must prioritize.

Across all stakeholder groups, some NFRs showed marked discrepancies between their simple Best-Worst "Net" scores and their model-based BWS coefficients. This pattern is expected given the properties of the exploded-logit model. The model estimates utilities conditional on the specific choice sets in which NFRs appear and relative to a reference item, so the resulting β values do not correspond linearly to raw Best/Worst counts. An NFR can accumulate many "worst" choices yet obtain a moderate β if those choices occur mainly in sets dominated by very strong competitors, whereas an NFR that wins primarily in comparatively easy sets may achieve a high Net score but only a modest model-based importance. For this reason, Net scores are interpreted descriptively, while the exploded-logit coefficients and importance shares provide the main basis for substantive inference about priorities for users, issuers, and verifiers.

Table 5.1: Overview of shared quadrant assignments across FI×BWS and FI×PI_{rev} for each stakeholder group.

Quadrant / Metric	**Identity Holders (Users)**	**Verifiers**	**Issuers**
Quadrant I (High–High)	Protection Security Transparency Control	Protection Authenticity Cost Privacy	Security Authenticity
Quadrant II (Low–High)	Cost Consent	Consent	Verifiability
Quadrant III (Low–Low)	Verifiability Existence Availability User Experience Interoperability	Accessibility Compatibility	Interoperability
Quadrant IV (High–Low)	Persistence	Verifiability	Cost Standard
Aligned NFRs / Total	12 / 24	8 / 13	6 / 12

A direct comparison of the FI×BWS and FI×PI_{rev} prioritization matrices (Table 5.1) shows that a sizeable subset of NFRs falls into the same quadrant across methods for all stakeholder groups (12/24 for users, 6/12 for issuers, 8/13 for verifiers). Core high-priority items (e.g., *Protection*, *Security*, *Transparency*, *Control* for users; *Authenticity* and *Security* for issuers; *Protection*, *Authenticity*, *Cost*, *Privacy* for verifiers) and several consistently low-priority NFRs are classified similarly, suggesting a stable core of priorities

that is robust to analytic choice. At the same time, the remaining NFRs shift across quadrants depending on whether importance is measured by abstract ratings or forced-choice trade-offs, underscoring that mid-tier requirements are more sensitive to framing and task format and should be interpreted with greater caution.

Prioritization matrices are presented separately for each stakeholder group rather than in a combined visualization because post-hoc pairwise comparisons revealed statistically significant differences in NFR ratings across roles for multiple quality attributes (Tables 4.48–4.49). These differences align with established findings that stakeholders prioritize quality attributes differently depending on their roles and responsibilities in software systems [36]. Separate matrices preserve role-specific prioritization patterns and ensure methodologically valid quadrant assignments using group-specific thresholds.

Findings have clear implications for requirements engineering. Where users demand qualities that providers undervalue, action is needed to prevent future dissatisfaction. For instance, users place *Transparency* in High FI–High PI, while issuers position it in Low FI–High PI, suggesting designers should add user-facing audit trails or communication mechanisms to bridge the divide. Conversely, provider priorities such as *Verifiability* may remain invisible to users but still require investment and user education about their importance for system trustworthiness. Resource allocation can be systematically guided by quadrant logic. All *role-specific* Quadrant I NFRs should be implemented early as non-negotiable foundations; the cross-role core comprises *Security*, *Protection*, and *Authenticity*, with *Privacy* joining Quadrant I for issuers and verifiers (users place *Privacy* in High FI–Low PI) and *Cost* appearing in Quadrant I for verifiers only. Quadrant II items (Low FI–High PI) warrant risk planning (e.g., users undervalue *Standards* until incompatibilities occur). Quadrant III items (Low FI–Low PI) can be deprioritized because they are neither important nor problematic at present; here, *Interoperability* sits for all three roles under the median split. Quadrant IV items (High FI–Low PI) represent qualities valued in principle but not currently problematic, which development teams can schedule flexibly once core requirements are satisfied. Statistical validation strengthens confidence: Wilcoxon tests confirmed Quadrant I items were rated significantly above neutral ($p < .05$), verifying genuine stakeholder consensus rather than measurement artifacts.

The prioritization matrices reveal not only which NFRs matter most to each stakeholder group but also why these priorities emerge from role-specific responsibilities. The quadrant approach distinguishes critical requirements from context-dependent ones, while methodological triangulation confirms the stability of priority rankings across evaluation methods. Combining FI and PI ratings, and BWS trade-offs provides a stakeholder-aligned view of NFR priorities. Statistical validation ensures high-priority requirements reflect genuine consensus rather than measurement artifacts. This multi-dimensional approach addresses pressing qualities for all stakeholders while identifying less obvious risks that simpler ranking methods might overlook, guiding more informed and balanced design decisions in SSI system development.

5.1.5 Cross-Role Differences in NFR Prioritization

While previous sections examined absolute priority rankings within stakeholder groups, pairwise post-hoc tests reveal significant differences in how roles assess specific quality requirements. These tests, conducted following significant Kruskal-Wallis results, identified the stakeholder pairs with the most pronounced divergences in priorities (see Tables 4.48 and 4.49).

Patterns in FI ratings showed that users consistently rated *Transparency*, *Compatibility*, and *Accessibility* significantly higher than organizational stakeholders, with large effect sizes indicating practically meaningful differences, extending [98] findings and revealing a notable gap between expert priorities and end-user perspectives. Users equate transparency with accountability and ease of use, suggesting these qualities primarily serve as trust-building mechanisms rather than technical requirements. For example, the study by [41] found that *Transparency* is commonly associated with positive concepts such as trust and accountability. Furthermore, the NFR *Accessibility* is recognized as an important quality that builds stakeholder trust in software systems.

Conversely, issuers and verifiers both prioritized *Verifiability* significantly higher than users, reflecting their operational need for reliable credential verification. This statement is also supported by [43], who state that, for identity systems to work effectively, identity credentials must be shared, fully validated, and verified in both directions to maintain trust. This means that verifiers and issuers need high-quality data and robust verification methods so credentials can be trusted across different systems [43]. This finding aligns with the survey results, which showed that verifiers and issuers rated verification as very important. The convergence among the actors within the organization, but the divergence vis-à-vis users, suggests that verification reliability is more important for issuers and verifiers than for users, as demonstrated by the study by experts [98].

Problem-Importance patterns revealed complementary, role-specific concerns. Issuers viewed *Verifiability* failures as significantly more severe than other stakeholders did, consistent with their existential concern that unverifiable credentials would undermine their entire operations [64]. Users and verifiers both rated *Standards* violations as significantly more problematic than issuers, suggesting that interoperability challenges manifest differently across roles [43, 98]. Users and verifiers, who interact with credentials across diverse platforms, directly experience standards fragmentation through failed verifications and incompatible formats [59]. Issuers, focused on their own credential issuance, may not recognize how standards violations create downstream problems for others [98]. Another observation is that *Security* failures were rated most severe by issuers, likely reflecting liability concerns where breaches could permanently damage their market position and reputation, leading to legal problems [10, 74]. This shows why issuers rank security failures as the most serious, as they face legal risks and brand damage.

The moderate-to-large effect sizes demonstrate that stakeholder disagreements represent fundamental divergences in operational priorities, not marginal differences in degree. Where no significant differences were observed, such as between issuers and verifiers across several dimensions, the roles shared similar concerns, indicating potential points of consensus. The findings validate stakeholder theory's prediction that role-based responsibilities

shape requirement prioritization [42]. However, this difference in transparency demonstrates that some requirements are valued mainly for their ability to build user trust and facilitate adoption, rather than for direct technical or operational reasons.

The cross-role differences also highlight limitations in expert-driven prioritization. [98] expert rankings systematically underestimated user priorities for transparency and compatibility, while potentially overestimating the universal importance of qualities such as standards [98]. This expert-practitioner gap suggests that requirement elicitation should involve actual stakeholders rather than relying solely on expert judgment. According to [73], RE practitioners struggle to balance the interests of various stakeholders, demonstrating that ignoring certain voices (for example, end-users) results in one-sided priorities. These findings confirm that designers cannot assume uniform NFR priorities across roles; interfaces and architectures must address role-specific concerns: Transparency for users, Verifiability for organizational actors, compatibility for users and verifiers, while maintaining baseline requirements that satisfy all groups [98].

5.2 Clarity of Quality Requirement Functionality

A key question is whether respondents clearly understood the linkage between system functionalities and underlying NFRs. The survey employed paired items: a FI and a PI statement for each NFR. The evidence suggests that clarity was mixed: users mostly understood the functionalities as intended, but specific items clearly needed rewording, and organizational stakeholders showed substantially less consistent interpretation.

For identity holders, FI and PI_{rev} rankings demonstrated a significant positive correlation (Spearman's $\rho = 0.561$, $p = .004$), indicating that users who rated a quality highly in general also found its failure scenario concerning. This group's Cronbach's alpha was high ($\alpha = 0.89$), suggesting strong internal consistency in their answers. For example, *Protection* received high ratings in both FI and PI contexts. Top priorities (*Protection*, *Security*, *Control*, *Transparency*) appeared in the upper-right quadrant of the FI-PI importance matrix, indicating a clear understanding. In contrast, verifiers and issuers exhibited substantially lower alpha values, especially in PI_{rev} items, implying inconsistent understanding or confusion about some items. In other words, users generally had a coherent interpretation of the NFR-function mappings, whereas verifiers and issuers showed less agreement across related items.

However, notable divergences likely reflect ambiguity in the questions. *Consent* showed the most striking discrepancy: Users' FI rating was relatively lower (M = 4.23, mid-ranked), yet the PI scenario ("misuse of personal data without permission") drew near-maximum concern (M = 4.38, rank 2), which suggests respondents did not recognize consent's importance until confronted with a violation scenario. Similarly, verifiers' *Consent* ratings jumped from 8^{th} in FI to 2^{nd} in PI, revealing latent importance. Some respondents may have misunderstood specific items in the survey. For example, the items of the NFR *Protection* may have been often conflated with general data privacy measures, whereas *Consent* (user permission) was interpreted variably; some treated it as a subset of privacy, others as distinct user control. Similarly, items in the NFR *Autonomy* may have been

ambiguous: Some participants may have interpreted it as system independence or control over one's own data, while others found the term abstract. The requirement *Accessibility* could also have been confusing: some could have interpreted it as physical or assistive access (e.g., access for users with disabilities) rather than availability or service access as intended.

These inconsistencies could be traced to problems with the survey instrument's question wording. Several items contained double-barreled questions or conceptually ambiguous phrasing [70]. The Autonomy problem scenario asked: "If I had to ask someone else to change or share my digital identity, it would be fine for me". This negative framing combined multiple concepts (asking permission, changing identity, sharing identity) within a single scenario. Respondents might focus differentially on permission requirements, change actions, or sharing actions, creating rating uncertainty. Users showed inconsistent Usability ratings: moderate FI agreement but variable PI concern, likely reflecting mixed implications across scenarios. *Accessibility* presented similar challenges. The FI item stated: "I need to access my identity data whenever I want". However, the PI item used the following phrasing (limited + acceptable): "Limited access to my identity data when I need it would be acceptable to me.", which could confuse respondents. Such constructs are known to produce ambiguous responses [48, 103].

The negative framing and reverse-coding of PI_{rev} items introduced additional cognitive complexity. While analytically appropriate (by ensuring that higher scores consistently indicate greater importance), this approach may confuse respondents. Participants might misread polarity or find it harder to gauge concern from negatively framed scenarios ("would be acceptable to me") than to rate positively framed features. However, to avoid response biases, half of the questions were framed in a positive direction and the other half in a negative direction. The dramatic difference in alpha values between users and organizational stakeholders indicates that mapping clarity varied considerably across items and stakeholder roles. Users had more items (24 vs. 12–13 for organizational roles), which may have improved scale reliability. However, statistical correlation patterns reveal opportunities for improving construct clarity.

In summary, specific quality requirements NFRs and their corresponding functionalities were not uniformly precise to all respondents. Double-barreled items and negative wording complexity align with lower FI–PI_{rev} alignment for affected items. However, top-ranked qualities showed consistency across FI, PI_{rev}, and BWS measurement approaches. For example, NFRs such as *Protection*, *Security*, *Cost*, *Authenticity*, and *Privacy* remained dominant priorities across all frames for each stakeholder role, suggesting the most salient requirements were communicated effectively. Clarity issues emerged primarily for mid- and lower-ranked items, demonstrating how subtle phrasing differences can significantly affect respondent comprehension. To improve future questionnaires, best practice is to ensure each question targets a single, clearly defined concept. Future research should avoid double-barreled items and double negatives, use simple, unambiguous language [70, 103], and pilot-test or cognitively pre-test the survey with representative respondents or experts to catch misinterpretations early [48, 103]. Breaking multi-faceted scenarios into focused, single-concept items would reduce interpretation variability and improve measurement precision.

5.3 Comparison with Existing Literature

Comparing stakeholder findings with the expert-based baseline from [98] reveals substantial consensus alongside notable divergences. [98] identified *Security*, *Protection*, *Verifiability*, and *Authenticity* as the highest-priority qualities among experts. This aligns with the present findings: all three stakeholder groups prioritized *Protection* and *Authenticity*, consistently ranking these qualities at or near the top in both FI and BWS. The convergence on *Security* and *Authenticity* as foundational requirements suggests that both experts and practitioners recognize trustworthy credentials as essential to SSI ecosystems. Multiple studies, including [44], show that experts and stakeholders rate *Security* and *Privacy* as top priorities for DI systems. A value-sensitive design study of credential apps, for example, found that stakeholders consistently saw privacy, security, and trust as the most important qualities [44]. Similarly, studies that reviewed requirements for blockchain identity systems list *Provability* (which matches *Authenticity* in this study, meaning verifiable authenticity) and *Security* as core system qualities [106].

Privacy showed similarly strong agreement across studies. While [98] rated *Privacy* as among the top properties, stakeholders demonstrated comparable prioritization, with identity holders (users), verifiers, and issuers assigning it high importance. In this master's thesis, *Control* was assessed only by identity holders (users), not by organizational actors. *Consent* emerged similarly across studies, with stakeholders strongly prioritizing it across all roles. In practice, user-controlled consent (where users can approve or refuse data sharing) is highly valued by users, reflecting their desire for control over their personal data, even if some experts prioritize other system qualities [44, 51]. User control and ownership also aligned, reflecting shared recognition of the importance of individual agency in credential management [51]. Research shows that end-users especially demand control over their identity data. User-centered wallet designs aim to empower users by enabling them to manage their own digital identities [51].

Usability, *User Experience*, and *Accessibility* were valued but assigned lower priority than *Security* and *Privacy* concerns. While SSI systems must be usable to achieve adoption, usability operates as an enabling quality rather than a fundamental requirement. Similarly, *Availability* and *Persistence* followed comparable patterns, valued but considered less critical than trust-related attributes. Features such as usability, accessibility, high availability, and easy recovery are generally considered important, especially by users (who often rate them highly in surveys). For example, according to [51], previous ID-wallet projects have often reported that poor usability led to low adoption. However, they receive lower priority when systems must choose between features. *Recoverability* presented an interesting case: stakeholders rated *Recoverability* as important in absolute terms, but when forced to make explicit trade-offs, other qualities took precedence, reflecting a classic security-usability trade-off where users prioritize security and privacy concerns over recovery mechanisms.

A significant divergence concerns *Decentralization* and *Autonomy*. [98] found this property contentious among experts, yet stakeholders displayed strikingly divergent patterns across roles. This contradictory pattern among organizational actors suggests profound internal disagreement or contextual sensitivity regarding the value of decentralization. While decentralization may be architecturally important, its perceived value varies dramatically

across roles, echoing broader adoption challenges in DI systems where user-facing benefits of decentralization remain poorly communicated [88]. Though conceptually related, the underlying architectural principle of autonomy from central authorities does not resonate as strongly as direct control over credentials. In short, stakeholders often see *Decentralization*, *Autonomy*, and *Interoperability* as less important than experts do, because users cannot clearly see their benefits [44, 88].

Standard and *Interoperability* showed mixed patterns. Experts [98] rated *Standard* highly and selected *Interoperability* frequently. Verifiers aligned with this view, rating *Standard* high and assigning *Interoperability* substantial weight when making explicit trade-offs, confirming its operational necessity. However, users gave *Standard* lower priority and minimal emphasis, while finding its absence moderately concerning, suggesting latent awareness without explicit prioritization. Issuers rated *Standard* high but viewed its absence as least problematic, reflecting focus on credential issuance over ecosystem-wide *Interoperability*. *Compatibility* with legacy systems received varied treatment across roles, reflecting organizational integration needs. *Portability* likewise received moderate priority, indicating shared but measured concern for credential mobility across systems.

Transparency emerged more prominently for stakeholders than expected from expert rankings. While [98] did not place *Transparency* in the top tier, users prioritized it highly and assigned it substantial importance when forced to choose among competing qualities, surpassing *Privacy* and *Consent*. Issuers gave *Transparency* lower FI but recognized a lack of transparency as a significant problem. Verifiers, however, assigned *Transparency* a minimal priority. Users' elevated concern for *Transparency* indicates it serves as a trust-building mechanism that may be more critical to adoption than expert assessments initially suggested. However, studies of stakeholders' preferences for SSI systems often mention *Transparency* alongside *Security* and *Privacy*. For example, research on SSI design identifies *Transparency* as something users frequently mention as important in real-world SSI systems [44]. In summary, users see *Transparency* (clear policies, visible data flows, the ability to check actions) as important for building trust, even though some experts do not emphasize it as much [26, 44]. In contrast, *Representation* and *Single Source* ranked low across perspectives, possibly because these concepts remain abstract without clear user-facing benefits.

Cost and *Security* warrant special attention due to their counterintuitive positioning. *Cost* received moderate FI ratings across roles but substantial importance when stakeholders were forced to make trade-offs, indicating that although *Cost* is not strongly emphasized in absolute terms, it becomes a major constraint during decision-making, particularly among organizational actors. In fact, [88] note that users will not adopt SSI systems unless the new technology offers better economic value than what they currently use. In other words, even if *Cost* is not the most important priority, it becomes a deciding factor when choices must be made, especially for organizations (such as verifiers or service providers) with limited budgets. This suggests divided expert opinions on whether financial considerations should influence identity system design. Importantly, comprehensive studies of requirements list *Cost* as a key system quality that affects identity holders, issuers, and verifiers [106]. This shows that affordability may seem less important in surveys but is actually a key concern when organizations choose systems [88, 106]. *Security*, conversely, received high FI ratings across all roles but showed divergent priorities when forced to choose.

These findings indicate that while stakeholders view security as functionally important in abstract terms, its practical relevance varies considerably by role.

Existence and *Verifiability* complete the comparative picture. For identity holders, *Existence* was viewed as a secondary responsibility and received low emphasis, likely because having an identity is generally assumed and does not require explicit recognition or prioritization. *Verifiability*, in contrast, shows strong role-specific differentiation: organizational actors prioritized it substantially when making explicit trade-offs, validating its criticality for the ecosystem even if individual users rank it lower when not considering the broader system context. In practice, verifiers actively demand this proof: Studies of SSI find that a service provider or verifier holds a stronger position and can determine what identity data is required, potentially forcing users to disclose more than they wish [88]. This reflects different priorities between the two groups: verifiers rank *Verifiability* highly (since they need to trust credentials). At the same time, users tend to prioritize *Privacy* and *Accessibility* more, which aligns with surveys showing users emphasize *Verifiability* less.

The comparison reveals strong consensus on foundational priorities: *Security*, *Protection*, *Authenticity*, *Privacy*, and *Control* are universally recognized as critical by both experts and stakeholders. This alignment indicates that industry efforts and user expectations converge on essential SSI requirements [88]. However, experts emphasize *Decentralization*, *Interoperability*, and *Standard* more strongly than most practitioners. Research indicates that experts often give more weight to system architecture qualities (for example, *Decentralization* and *Interoperability*) than to what matters most to users right now. *Interoperability*, for example, is listed as a top NFR in many technical frameworks (e.g., [98, 106]), and *Decentralization* is a foundational principle in SSI standards. However, as mentioned above, stakeholders (especially end-users) may not see these features as important unless they understand why they matter. The gap between what experts focus on and what users want has been noted: features such as *Decentralization* and *Interoperability* are widely discussed by designers and researchers, but were largely absent from early studies of what users and stakeholders value [44, 106]. It is suggested that expert assessments lean toward technical ideals (such as decentralized governance and open standards) even when these are not among users' most important features.

Researchers note that forced-choice methods, such as BWS, can reveal priority differences that simple rating scales may hide. BWS asks respondents to choose the "most" and "least" important items in each set, which makes differences in preferences clearer. In fact, BWS was designed to replace traditional rating scales because these cannot accurately predict choices and have several problems [71]. In SSI systems, this means certain system qualities (especially those users see as important for everyone, like *Usability* or *Accessibility*) tend to receive high scores, which hides which ones really matter most. When users must choose between items (as in BWS), differences emerge. This methodological advantage explains why BWS often reveals hidden priorities (like cost or specific trust qualities) that simple surveys miss [71]. These findings empirically ground the expert-derived property set with stakeholder perspectives, confirming that non-negotiable requirements include *Security*, *Protection*, *Authenticity*, *Privacy*, and *Control*, while *Transparency* and role-specific operational concerns warrant greater design attention [10, 21, 64, 76].

5.4 Construct Validity and Reliability of the Findings

The internal reliability of the measurement scales varied substantially across stakeholder groups, with important implications for how the findings should be interpreted. Cronbach's α measures the extent to which survey items consistently assess the same underlying construct, with values above 0.70 considered acceptable and above 0.80 indicating good reliability [87]. In practice, an alpha value around 0.7 indicates acceptable reliability, while values of 0.8 or higher indicate stronger internal consistency [14]. Users demonstrated strong internal consistency, while verifiers and issuers showed weaker measurement reliability.

For identity holders (users), both the FI and PI_{rev} scales showed high internal consistency, with Cronbach's α's of 0.890 and 0.889, respectively, exceeding the 0.80 threshold for good reliability. This indicates that users interpreted the quality requirements consistently and understood the distinctions among the NFRs. The α-if-deleted analysis revealed that removing any single item had minimal effect on overall reliability, suggesting that all items contributed meaningfully to the scales. Only *Representation* showed a slight potential improvement upon removal. This strong reliability supports the validity of the user-priority findings and confirms that the functional descriptions effectively communicated distinct quality concepts to non-technical stakeholders.

For verifiers, the FI scale demonstrated acceptable reliability ($\alpha = 0.762$), meeting the minimum threshold of 0.70, but the PI_{rev} scale showed substantially lower consistency ($\alpha = 0.445$), indicating questionable reliability. Verifiers either interpreted the problem scenarios inconsistently or experienced genuinely different challenges in their verification contexts. The α-if-deleted diagnostics revealed that removing *Interoperability* would improve both scales, indicating this item was less aligned with the others. The low PI_{rev} reliability suggests that verifiers may have more diverse operational experiences than users, reflecting differences in verification use cases, technical maturity, or organizational contexts rather than measurement error.

For issuers, both scales showed low internal consistency, with α values of 0.584 for FI and 0.472 for PI_{rev}, falling well below conventional reliability thresholds. This suggests that issuers interpret quality requirements differently or that the construct is multidimensional. The low reliability among issuers likely reflects genuine diversity in organizational priorities rather than poor question design, as different types of issuing organizations face distinct challenges depending on their industry, size, and regulatory context.

The contrast between strong user reliability and weak organizational reliability has important implications for interpreting the findings. High user reliability suggests that individual end-users have relatively uniform expectations for SSI systems, making their collective priorities easier to identify and address in system design. In contrast, low organizational reliability suggests that verifiers and issuers have more heterogeneous needs, meaning that one-size-fits-all solutions may not satisfy all types of organizations. This finding supports the need for flexible, customizable SSI implementations that can adapt to different organizational contexts rather than rigid, uniform architectures.

The inter-item correlations provide additional evidence for construct validity by showing which NFRs stakeholders perceive as related. For users, strong correlations in the FI block

emerged between *Compatibility* and *Portability*, *Compatibility* and *Usability*, *Authenticity* and *Portability*, *Compatibility* and *Security*, and *Authenticity* and *Compatibility*, reflecting logical clustering of user-facing technical requirements. These patterns suggest that users perceive systems supporting credential portability as inherently more secure and authentic [64]. The correlation between *Recoverability* and *Security* indicates that users associate account recovery mechanisms with overall system security.

These correlations raise questions about whether users distinguish between distinct quality dimensions or conflate multiple attributes into broader categories. *Cost* showed strong positive correlations with *Standards*, *User Experience*, and *Portability*, indicating that users perceive affordable systems as also being technically robust and easy to use, not as a trade-off [98]. This pattern may reflect that users expect efficient, well-designed SSI systems to integrate *Portability*, *Standard*, and *Usability* as complementary rather than competing attributes.

The NFR *Representation* showed consistent negative correlations with multiple quality dimensions in the FI block (*Compatibility*, *Security*, *Verifiability*), suggesting that these quality dimensions may have been ambiguously worded. This pattern indicates a measurement validity issue warranting revision in future studies. In practice, NFRs are often poorly defined and difficult to verify, leading respondents to interpret them differently [101]. NFRs may also naturally conflict, overlap, or complement one another, making it difficult for users to distinguish among them [4]. The consistent negative correlations involving *Representation* provide strong evidence for either a double-barreled question formulation or a conceptual mismatch between the functional description and the underlying quality attribute [67].

For the PI_{rev} block, the strongest correlation was between *Compatibility* and *Interoperability*, suggesting users perceive these as conceptually related. According to [69], *Compatibility* and *Interoperability* form one quality category concerning how different services work together and include features such as openness and reusability. Other notable positive correlations included *Compatibility* and *Standards*, *Compatibility* and *Portability*, *Authenticity* and *Standards* compliance, *Interoperability* and *Verifiability*, and *Security* and *Standards*. Standards compliance influences other qualities, such as the ease with which a service can be moved and used across systems (*Portability* and *Interoperability*) [69]. These patterns reflect that users' concerns about implementation problems cluster around technical integration and regulatory adherence. The consistently high correlations centered on *Compatibility* suggest that users perceive multiple interoperability-related quality dimensions as interconnected rather than independent concerns [104].

For verifiers, strong correlations in the FI block emerged between *Transparency* and *Security* and *Privacy* and *Standards*, reflecting that verifiers perceive secure verification processes and regulatory compliance as interconnected operational requirements. A slight negative correlation between *Transparency* and *Decentralization* suggests that some verifiers perceive these as competing architectural priorities. For the PI_{rev} block, the strongest correlation was between *Privacy* and *Authenticity*, though correlations remained largely non-significant, indicating that verifiers experience implementation problems heterogeneously across their specific verification contexts.

For issuers, the strongest correlation in the FI block was between *Protection* and *Security*, reflecting the conceptual overlap between these security-oriented requirements in credential issuance [98]. For the PI_{rev} block, the strongest correlations emerged between *Authenticity* and *Verifiability* and *Protection* and *Verifiability*, indicating that issuers experience security-related problems together in current systems. The low percentage of significant correlations across organizational stakeholders confirms that the survey successfully measured distinct quality dimensions while revealing more context-dependent needs than individual users.

Lower Cronbach's α values for verifiers and issuers reflect genuine differences in how stakeholder groups interpret quality requirements rather than survey design flaws. Cronbach's α appropriately varies across different groups and contexts, as α depends not only on survey questions but also on respondent consistency and operational situational factors [87]. Subtle differences in situation can alter how respondents interpret items and their relationships with others. For verifiers and issuers with diverse operational environments across verification use cases and organizational types, lower α values are methodologically appropriate and indicate fundamental role-based divergence in the interpretation of quality requirements.

5.5 Implications for SSI Design

Different stakeholders prioritize different quality requirements based on their roles in SSI ecosystems. Users prioritize *Protection*, *Security*, *Control*, and *Transparency*, indicating that SSI wallets must prioritize data safety and give users control over their information [93, 98]. Practical design needs include strong encryption, detailed consent controls, clear audit trails, and reliable recovery methods. Users fear losing permanent access to their digital identities, which prevents adoption despite other advantages [84]. SSI wallets must therefore offer multiple recovery methods (e.g., social recovery, encrypted backups) while maintaining strong security to build user confidence [25, 84].

Verifiers emphasize *Authenticity*, *Cost-Efficiency*, and *Interoperability*. The divergence between verifiers' stated importance of qualities and their actual trade-off choices shows that *Cost-Efficiency* is a critical adoption barrier. SSI reduces verification costs by automating checks, eliminating manual paperwork, and reducing human errors and fraud risk [6]. SSI solutions must prioritize low-cost verification processes through efficient protocols, cached credentials, or standardized verification APIs [88]. Verifiers also view standards compliance as something developers should handle in the background. Systems should, by default, integrate with existing infrastructure using widely adopted standards such as W3C Verifiable Credentials to ensure interoperability without requiring explicit verifier requests [59, 64].

Issuers prioritize *Security*, *Protection*, and *Verifiability*, requiring systems that guarantee credential integrity and reliable verification even when issuers are offline [88]. After issuing a verifiable credential with a digital signature, holders can present credentials to verifiers without contacting the issuer [6]. Quality requirements should be mapped to the parties

responsible for them, with issuers defining authenticity specifications, verifiers specifying verification workflows, and users determining control and transparency requirements [98]. *Transparency* emerged as more important to users than expert rankings suggested, functioning as a trust-building mechanism rather than merely a technical feature. SSI systems should provide visible audit trails, clear consent logs, and understandable credential lifecycles to address users' transparency needs [25, 64, 93].

Successful SSI implementations require customization for each role rather than a one-size-fits-all design. *Security*, *Protection*, *Authenticity*, *Privacy*, and *Control* emerge as baseline requirements that must be satisfied across all stakeholder groups, confirming expert-based priorities identified by [98]. Additional features should reflect specific role needs: user-facing interfaces must emphasize *Transparency* and *Recoverability*, verifier systems must optimize for *Cost* and *Interoperability*, and issuer infrastructure must ensure *Verifiability* and compliance with *Standards* [10, 64]. Each stakeholder group benefits differently: SSI issuers gain efficiency and reduced risk, verifiers reduce costs while building trust, and users gain control over their data and privacy [6]. Users showed limited interest in decentralized architecture itself, suggesting that SSI systems should prioritize communicating user-facing benefits, such as *Control*, *Privacy*, and *Security*, rather than technical architecture details [88]. Designers must emphasize *Decentralization*'s practical advantages to drive user adoption.

5.6 Implications for Requirements Engineering Practice

Requirements engineering for multi-stakeholder systems, such as SSI, requires new approaches to capture each group's distinct priorities. The combination of FI ratings, PI rankings, and BWS yields distinct insights into what stakeholders truly value. Relying on a single measurement method can miss important information [54]. Multiple measurement techniques provide a better understanding of diverse priorities than single methods in complex systems where stakeholders have different goals [35, 57]. FI ratings capture what stakeholders think matters in general, while BWS forces trade-offs and reveals what they actually prioritize when they cannot have everything [89]. Combining multiple methods, such as importance ratings and BWS, provides a more complete picture of stakeholder needs, ensures that role-specific priorities are correctly understood, and helps balance requirements in final specifications [29].

The findings show that stated importance and forced trade-offs yield different results, with practical implications for requirements gathering and prioritization. BWS reveals true priorities by asking respondents to choose between "most" and "least" important items, whereas simple rating scales often yield many items rated highly important [29]. Requirements engineers should use prioritization techniques like BWS to identify which requirements are truly non-negotiable versus which ones stakeholders would sacrifice when faced with resource limits [3, 83]. This approach prevents teams from overinvesting in features that stakeholders claim are important but would actually compromise to reduce costs or improve other qualities. According to [98], SSI systems cannot achieve all desired properties simultaneously, so identifying core properties to prioritize is essential. Trade-off methods like BWS help stakeholders distinguish requirements they consider necessary

from those they find desirable only in theory, focusing work on requirements that genuinely cannot be compromised under real conditions.

Gathering requirements based on roles emerges as a necessary practice for SSI and similar multi-stakeholder environments [10]. Requirements should be mapped to stakeholders directly responsible for implementing or managing each quality dimension: users should define *Transparency* and *Control* requirements, issuers should specify *Authenticity* and *Verifiability* standards, and verifiers should design verification workflows [98]. For example, *Control* and *Recoverability* matter most to users who manage and recover their identity information. In contrast, verifiers focus on *Authenticity* and *Security* to verify the validity and source of credentials. This approach ensures that technical specifications come from parties who understand real-world context best, reducing the risk of defining requirements that sound good in theory but prove challenging to implement [10]. In e-government projects, meeting role-specific needs through focused requirements practices leads to higher stakeholder satisfaction, whereas failing to address them results in misalignment and user frustration [5]. Organizations should actively gather and organize requirements from each stakeholder group, mapping system features to responsible stakeholders, which simplifies implementation and creates solutions that each group finds acceptable.

Experts and practitioners often prioritize different things than actual users and stakeholders. Requirements experts consider important (such as strong security, strict compliance with standards, or new technical features) may not align with what users truly need. In digital identity systems, expert frameworks typically focus on maximum privacy or strong security. However, studies find that regular users place higher value on practical qualities such as *Transparency*, *Usability*, *Recoverability* of access, and *Accountability* [77]. This master's thesis shows that requirements engineers must consider the gap between expert opinions and stakeholder needs. While experts correctly identified *Security*, *Protection*, and *Privacy* as foundational, they underestimated the extent to which users value *Transparency* and *Recoverability*. This suggests that expert-based requirement prioritization should be validated with actual stakeholders before implementation, particularly for qualities affecting trust and adoption [88]. According to [5], insufficient stakeholder involvement early on leads to requirements that miss user needs, resulting in poor adoption and potential failure. To avoid this, requirements engineering should be an ongoing, human-centered process in which expert assumptions are tested against stakeholder needs through validation phases, ensuring that expert-derived priorities align with the real concerns of identity holders (users), issuers, and verifiers who will use the system. This approach ensures final requirements align with the fundamental values and concerns of people who will depend on the system.

Chapter 6

Final Considerations

The final chapter checks the research outcomes against the original aims and looks back on the project's process. It assesses whether the set objectives were met, summarizes the main results and their impact, and notes any difficulties or changes encountered during the work. It ends with a discussion of the study's limitations and ideas for future research.

6.1 Summary

The main goal of this study was to determine which NFRs are most important to different stakeholder groups in SSI ecosystems and to assess whether participants clearly understood how functional scenarios were linked to these requirements. The research used a mixed-methods approach, including a detailed questionnaire and several analytical techniques. The questionnaire collected two types of Likert-scale ratings (SQRI, inspired by the Kano model): FI and PI for each quality. Participants also completed BWS tasks. Respondents were divided into three main SSI stakeholder groups: identity holders (users), credential verifiers, and credential issuers, including 86 users, about 27 verifiers, and 37 issuers. The study used both quantitative and qualitative analysis to answer two questions: (1) which qualities each stakeholder group considered most important, and (2) if stakeholders understood the mapping from survey scenarios to the NFRs as intended.

Objective attainment: Overall, the project reached its main objectives. For the first research question, distinct priority profiles for stakeholder groups were found using prioritization matrices. The results showed which NFRs each group found most important and highlighted both shared and different priorities among identity holders (users), verifiers, and issuers. For the second question, the clarity of mapping between functions and NFRs was assessed. Most participants, especially end-users, understood the survey items, but some misunderstandings occurred, mainly among organizational roles. Thus, the mapping clarity objective was only partly achieved. The data revealed areas of clear and unclear understanding. The main reason was the complex question wording, which was identified as a limitation. Nonetheless, meaningful conclusions and areas for improvement in the questionnaire were identified.

Research process summary: The project moved through several phases to reach these results. It started with a thorough literature review and categorized NFRs, creating a baseline of expected qualities for each role, using expert sources such as [98] on SSI quality frameworks. Based on this baseline, a survey was developed to cover all key qualities and be clear to participants. Running the survey was challenging but successful. The project collected responses from all target groups, though it took extra effort to recruit credential issuers and verifiers, as their roles are specialized. Data analysis included descriptive statistics, correlation analysis, and visual tools such as prioritization matrices, allowing for results to be viewed from different perspectives. During the process, minor changes were made to keep the study on track. Overall, the project met its aim by providing empirical insights into stakeholder priorities and by evaluating participants' understanding, helping fill the gap in user-centered requirements prioritization for SSI.

6.2 Conclusions

This research offers several important conclusions about the priorities of quality requirements in DI applications and SSI ecosystems. The findings show that stakeholder roles have distinct priorities, reflecting the responsibilities and concerns of each role. However, there is also a core group of qualities that all roles value highly, highlighting universal needs in SSI systems.

- **Stakeholder priority profiles:**

  Each stakeholder category prioritized qualities aligned with their role in the ecosystem. Identity holders (users) placed the most significant importance on qualities directly affecting their personal *Security*, *Privacy*, and *User Experience*. They consistently rated *Protection* of identity data, *Security* of the system, *Recoverability*, *Control* over personal data sharing, and *Transparency* of the system's operations among the top requirements. In contrast, users showed relatively less concern for more abstract or system-centric qualities, such as total *Decentralization* or *Autonomy* from third parties. These principles, while foundational to SSI in theory, were not seen as immediately critical to users compared to concrete *Security* and *Usability* needs.

  Credential verifiers (e.g., organizations that verify credentials) placed greater emphasis on different aspects. Verifiers highly prioritize qualities that ensure the trustworthiness and efficiency of the verification process. For instance, *Privacy* of user data emerged as a top concern even for verifiers, likely because maintaining user trust and compliance is important for them. They also valued *Authenticity* of credentials and *Protection* against fraud, as well as adherence to *Standards/Interoperability* (since standardization eases verification across systems). The NFRs *Cost* and *Security* of the verification process were also important to verifiers. However, verifiers generally gave lower priority to user-centric qualities such as *Transparency* and *Accessibility*. These aspects matter to users but are less directly relevant to their operational roles.

  Credential issuers showed yet another profile, though with some overlap. Their priorities centered on the *Security* and reliability of the credentials they issue and the

infrastructure supporting them. Issuers highly valued *Security* measures, *Protection* of credentials, and *Authenticity/Verifiability*, ensuring that credentials are trustworthy and cannot be easily faked. They also appreciated *Standards* compliance, which facilitates broad acceptance of their issued credentials. Qualities related to system performance and integrity were more prominent for issuers, whereas, similar to verifiers, they gave comparatively low importance to end-user experience qualities (for example, user *Control* or *Transparency* might not have been top-of-mind for issuers in the results). In summary, users prioritize personal data protection and usability; verifiers focus on trust, authenticity, and efficiency; issuers emphasize security and trustworthiness of issuance. This differentiation supports the initial hypothesis that stakeholder roles inherently shape which qualities are deemed most important.

- **Consensus on critical qualities:**

  Although stakeholder groups had different priorities, all agreed on some fundamental NFRs. *Security*, *Protection*, and *Privacy* were consistently top priorities in both SQRI ratings and BWS tasks across all groups. *Authenticity* was another essential requirement for all stakeholders. This broad agreement shows that *Security*, *Privacy*, and *Authenticity* are non-negotiable requirements for SSI systems. Our results align with what experts in the field [98] have emphasized, confirming that *Security* and *Verifiability* are the most important. The stakeholder data from our study supports the expert consensus.

  However, the study also exposed gaps between stakeholder expectations and SSI theory. Principles such as *Decentralization* and *Autonomy*, often viewed as essential by SSI architects, did not rank highly among real users in this survey. Many users did not care whether the system was decentralized or governed by an authority, as long as their concerns about *Security*, *Privacy*, and *Usability* were addressed, suggesting that users may not fully appreciate or value certain key design principles, possibly because these qualities are hard to see in everyday use. SSI developers should not assume users share expert priorities. Instead, they may need to communicate the benefits better or keep such principles in the background, so users get what they want while systems maintain expert standards.

- **Clarity of NFR-functional mapping:**

  Another main finding is how well survey participants understood the link between described functionalities and quality requirements. Understanding was mixed; stakeholder groups and NFRs showed distinct patterns. For identity holders (users), understanding was strong: their responses to abstract importance ratings and scenario questions were consistent, indicating they interpreted each quality similarly. Statistically, users showed high internal consistency, with a strong relationship between how important they rated a quality and how concerned they were about its potential failure. The mapping between functional scenarios and NFR concepts was precise.

  On the other hand, verifiers and issuers showed less consistency in their responses, suggesting that some NFR items were less clear or relevant to them. Lower reliability scores and weaker links between FI and PI ratings suggest some survey questions were confusing or interpreted differently. Specific terms, such as *Consent* and *Autonomy*, were often understood differently. For example, many participants

did not rate *Consent* as important until confronted with a scenario about a violation, making its importance more obvious. Similarly, double-barreled or negated wording likely led to misunderstanding; for instance, a question combining *Control* with data sharing may have confused issuers. These results indicate that the clarity of NFR-functional mapping was only partial, with most confusion arising from the survey's design.

The implications are twofold. First, the main conclusions about stakeholder priorities are trustworthy because the highest-ranked qualities were understood clearly by all groups. Most of the confusion occurred with items that were mid-ranked or less obviously important. Second, the results show that questionnaire design is crucial when studying abstract concepts such as software qualities. Unclear wording can strongly affect how people interpret and rate items. In this study, this means the top priorities are reliable, but differences among lower-ranked items should be interpreted with caution due to potential measurement noise. Still, finding the confusing items is useful, as it shows how future surveys can be improved.

- **Reflection on methodology:**

  The study's multi-method approach highlights its advantages in revealing stakeholder priorities. A simple rating scale alone would not have shown much detail. Most qualities received high ratings, creating a ceiling effect and making it hard to determine which was most important. Including BWS required participants to make hard choices, which clarified which qualities mattered most. Comparing Functional and Problem Importance also revealed hidden concerns. Visual matrices helped show where abstract ratings and practical ratings aligned or differed within each group. These techniques strengthened our results by confirming priorities across methods. Disagreements between methods signal where more research is needed. The main lesson is that a multi-method approach provides a better, more complete view of stakeholder needs.

  Overall, this master's thesis met its objectives. It confirmed expected priorities in SSI systems. For example, *Security* and *Privacy* are key for all roles, and new insights include differences driven by stakeholder roles and the challenge of making abstract requirements clear to diverse groups. The findings highlight the need to design identity systems tailored to the specific needs of identity holders (users), verifiers, and issuers, and underscore the importance of question design for gathering precise requirements from end-users. The conclusions are helpful both for the SSI sector, by clarifying what matters most to different actors, and for requirements engineering, by showing how a user-focused approach works and what difficulties may arise.

6.3 Limitations and Future Work

While this thesis addressed its research questions, it is important to note its limitations and boundaries. These limitations help put the results in context and suggest areas for further study. Additionally, the findings point to several future research directions that could expand on the insights from this work.

6.3.1 Limitations

One limitation of this study concerns the sample of participants. Although a decent number of identity holders (users) were surveyed, the groups of credential verifiers and issuers were relatively small (in the tens), which could skew the importance ratings, as highly engaged users might prioritize differently than a truly general population.

Another limitation is the geographical and organizational diversity of respondents: the study did not specifically control for or segment results by region or industry. Cultural factors or the specific context in which a verifier or issuer operates (e.g., the financial sector vs. government) could influence which qualities they prioritize. Our results represent an aggregate view and might not capture these nuances.

There are also limitations in the research instrument and methodology. As discussed, several survey questions proved confusing or double-barreled, especially for non-user stakeholders, which may explain lower consistency in those groups' responses. In hindsight, the survey's specifically formulated NFR items/statements may not have been conveyed in the simplest possible terms, potentially leading to misinterpretation.

Furthermore, issuers and verifiers were given a shorter survey focused solely on the qualities they considered most relevant. The shorter survey length helped avoid participant fatigue but limited direct comparability, where users rated 24 qualities, while the other groups rated about half as many. However, from the start, only the NFRs relevant to each stakeholder were included, selected during the NFR mapping phase at the beginning of the study. One advantage of the survey design was that, in the SQRI Likert scale tasks, the items for which each stakeholder had primary responsibility were presented first. Sections on secondary and tertiary items followed these. The NFRs with primary responsibility were prioritized in the survey.

Finally, self-reported importance ratings are always subjective. What participants say is important may not match how they actually behave in real-life situations. Although the study included scenario-based questions and BWS tasks to reduce this effect, it still measures what people claim to prefer rather than what they actually do, which is a general limitation of survey research.

Despite these limitations, the study's main findings are robust within the context of the sample and methods used. The identified limitations mainly highlight areas where improvements or more data could strengthen the results. They also provide guidance for designing future research on this topic.

6.3.2 Future Work

Building on this thesis, there are several promising directions for future research and development:

Future research should improve questionnaire design by removing ambiguity and making statements more transparent and more focused on single qualities. Negative or complex phrasing should be avoided to reduce confusion. It is recommended to pilot-test any revised survey with a small group of stakeholders to identify and fix unclear items before wider use. Additionally, future studies should include larger, more diverse groups of verifiers and issuers, and possibly other stakeholders such as SSI platform providers or relying parties, to determine whether the observed priority patterns apply more broadly. With a larger sample, it would also be possible to analyze differences between subgroups, such as comparing priorities among financial industry verifiers and government verifiers.

Future studies should include qualitative research to deepen understanding of stakeholder priorities. Interviews or focus groups with representatives from each stakeholder group could explore questions like: Why do end-users not care much about *Decentralization*? What experiences make verifiers highly concerned about *Privacy*? How do issuers balance *Security* and *Protection*? These insights would complement the survey's quantitative results by explaining the reasoning behind the rankings. Qualitative methods could also reveal misconceptions or hidden requirements that may not be identified through structured surveys.

Longitudinal and iterative research would add value by recognizing that requirements prioritization in SSI may change over time. Longitudinal studies could track how stakeholder priorities shift as SSI technology develops and becomes more common. For example, if users learn more about SSI or significant security incidents occur, their views on qualities such as *Autonomy* or *Transparency* may change. Conducting similar surveys repeatedly over several years, or after significant events such as new laws or major SSI rollouts, could reveal how priorities evolve.

Broader application and generalization: The methods and approach used in this thesis can be applied to other areas beyond SSI to study user-centered requirements prioritization. Future research could examine emerging technologies such as digital health records, smart city platforms, and blockchain applications to determine whether different stakeholders show similar patterns in the qualities they prioritize. Comparing findings across different domains will help researchers see which results are specific to SSI and which are universal principles of stakeholder prioritization. For SSI itself, future studies could include additional roles, such as governance authorities or trust framework providers, to create a more complete picture of the ecosystem's requirements.

Design implications and tool development: The findings from this study can inform the design of SSI systems and their development guidelines. Future research could turn these priorities into practical design requirements or best practices for SSI developers. For example, since users value *Recoverability* and *Control*, developers might add new wallet features to improve account recovery and more granular consent management. Understanding verifier concerns about *Standards* and *Cost* could help improve protocol standardization and

system efficiency. Researchers and SSI platform developers could work together to make these changes and then study whether they lead to higher satisfaction or wider system adoption. A valuable direction is to use these insights to build better systems and then test how well they meet stakeholder needs.

Reconciling stakeholder differences: The study revealed specific misalignments, such as qualities that users might overlook but are important for verifiers, and vice versa. Future research could focus on ways to balance or reconcile these differences. One possible approach is to develop negotiation frameworks or tools in requirements engineering to help different stakeholder groups work together to set priorities. Another is to explore role-specific responsibility, where each critical quality is assigned to the stakeholder best able to ensure it. For example, if users do not prioritize *Interoperability*, issuers and verifiers could take responsibility through standards development, ensuring that important qualities are not ignored. Research could also examine strategies for education and communication, helping users understand the value of lower-ranked qualities (such as *Decentralization*) and possibly increasing their awareness and concern for these aspects.

Bibliography

[1] ISO 25000. *ISO 25010*. URL: https://iso25000.com/index.php/en/iso-25000-standards/iso-25010.

[2] Will Abramson, Nicky Hickman, and Nick Spencer. "Evaluating trust assurance in Indy-Based identity networks using public ledger data". In: *Frontiers in Blockchain* 4 (Apr. 2021).

[3] Philip Achimugu et al. "A systematic literature review of software requirements prioritization research". In: *Information and Software Technology* 56.6 (Feb. 2014), pp. 568–585.

[4] Abdelkareem M. Alashqar. "Studying the commonalities, mappings and relationships between non-functional requirements using machine learning". In: *Science of Computer Programming* 218 (Mar. 2022), p. 102806.

[5] Asaad Alzayed. "Evaluating the role of Requirements engineering practices in the sustainability of electronic government Solutions". In: *Sustainability* 16.1 (Jan. 2024), p. 433.

[6] Matthias Babel et al. "Self-sovereign identity and digital wallets". In: *Electronic Markets* 35.1 (Apr. 2025).

[7] Yirui Bai et al. "Decentralized and Self-Sovereign Identity in the Era of Blockchain: A Survey". In: *2022 IEEE International Conference on Blockchain (Blockchain)* (Aug. 2022), pp. 500–507.

[8] Anum Bakhtiar et al. "PRIORITIZATION OF VALUE BASED SERVICES OF SOFTWARE BY USING AHP AND FUZZY KANO MODEL". In: *International Conference on Computational and Social Sciences* (May 2015).

[9] Pritha Bhandari. *Questionnaire Design | Methods, Question Types & Examples*. July 2021. URL: https://www.scribbr.com/methodology/questionnaire/.

[10] Ricardo Bochnia, Daniel Richter, and Jürgen Anke. "Self-Sovereign Identity for Organizations: Requirements for Enterprise software". In: *IEEE Access* 12 (Jan. 2024), pp. 7637–7660.

[11] Christoph H.-j. Braun et al. "SSI, from Specifications to Protocol? Formally Verify Security!" In: *Proceedings of the ACM Web Conference 2022* (May 2024), pp. 1620–1631.

[12] John Brooke and Usability Professionals' Association. *SUS: A retrospective*. Tech. rep. 2. Feb. 2013, pp. 29–40.

[13] Jan Camenisch and Anna Lysyanskaya. *An Efficient System for Non-transferable Anonymous Credentials with Optional Anonymity Revocation*. Jan. 2001, pp. 93–118.

[14] Gordon W. Cheung et al. “Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations”. In: *Asia Pacific Journal of Management* 41.2 (Jan. 2023), pp. 745–783.

[15] Kei Long Cheung et al. “Comparison of statistical analysis methods for object case best-worst scaling”. In: *Journal of Medical Economics* 22.6 (Nov. 2018), pp. 509–515.

[16] Fábio Avigo De Castro Pinto, Anarosa Alves Franco Brandão, and Fábio Levy Siqueira. “Design Thinking and Non-Functional Requirements Elicitation: A Survey”. In: *Anais do Workshop em Engenharia de Requisitos - Proceedings of the 25th Workshop on Requirements Engineering (WER2022)* (Jan. 2022).

[17] *Decentralized Identifiers (DIDs) v1.0*. July 2022. URL: `https://www.w3.org/TR/did-1.0/`.

[18] Omar Dib and Khalifa Toumi. “Decentralized Identity Systems: architecture, challenges, solutions and future directions”. In: *Annals of Emerging Technologies in Computing* 4.5 (Dec. 2020), pp. 19–40.

[19] Don A. Dillman, Jolene D. Smyth, and Leah Melani Christian. *Internet, phone, mail, and Mixed-Mode surveys*. Aug. 2014.

[20] Yepeng Ding and Hiroyuki Sato. “Self-Sovereign Identity as a Service: Architecture in Practice”. In: *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)* (June 2022), pp. 1536–1543.

[21] Dustin Doege, Ricardo Bochnia, and Jürgen Anke. *Fulfilling Principles of Self-Sovereign Identity: Towards a conformity Assessment Approach for human wallets*. 2024. URL: `https://dl.gi.de/handle/20.500.12116/44096`.

[22] Paul Dunphy and Fabien A.P. Petitcolas. “A first look at identity management schemes on the blockchain”. In: *IEEE Security & Privacy* 16.4 (July 2018), pp. 20–29.

[23] Serge Egelman, Marian Harbach, and Eyal Peer. “Behavior Ever Follows Intention?” In: *CHI ’16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (May 2016).

[24] Serge Egelman and Eyal Peer. “Scaling the Security Wall”. In: *CHI ’15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Apr. 2015), pp. 2873–2882.

[25] Edona Fasllija, Jakob Heher, and Stefan More. *Credential issuance Transparency: A Privacy-Preserving Audit Log of Credential issuance*. Jan. 2025, pp. 107–126.

[26] Simon Feulner et al. “Self-sovereign identity in the public sector: Affordances, experimentation, and actualization”. In: *Government Information Quarterly* 42.3 (June 2025), p. 102052.

[27] Kraig Finstad. “The usability metric for user experience”. In: *Interacting with Computers* 22.5 (Apr. 2010), pp. 323–327.

[28] Andrea Flamini et al. “On cryptographic mechanisms for the selective disclosure of verifiable credentials”. In: *Journal of Information Security and Applications* 83 (May 2024), p. 103789.

[29] Terry N. Flynn et al. “Best-worst scaling: What it can do for health care research and how to do it”. In: *Journal of Health Economics* 26.1 (May 2006), pp. 171–189.

[30] Floyd J Fowler Jr. *Survey research methods*. 5th ed. Sept. 2013.

[31] Sandro Rodriguez Garzon et al. “DID Link: Authentication in TLS with Decentralized Identifiers and Verifiable Credentials”. In: *2024 21st Annual International Conference on Privacy, Security and Trust (PST)* (Aug. 2Next), pp. 1–11.

[32] Aviral Goel and Yogachandran Rahulamathavan. “A Comparative survey of Centralised and decentralised Identity Management Systems: Analysing scalability, security, and feasibility”. In: *Future Internet* 17.1 (Dec. 2024), p. 1.

[33] Zoe Grotophorst et al. *The Differences in Data Cleaning Procedures in Probability and Nonprobability Panels*. Tech. rep. 2023.

[34] Spencer E. Harpe. “How to analyze Likert and other rating scale data”. In: *Currents in Pharmacy Teaching and Learning* 7.6 (Oct. 2015), pp. 836–850.

[35] Rohayanti Hassan et al. “Analysis of Multi-Stakeholder Requirements using Requirement interaction Matrix”. In: *International Journal on Advanced Science Engineering and Information Technology* 7.4-2 (Sept. 2017), p. 1498.

[36] Kennet Henningsson et al. “Understanding the Relations between Software Quality Attributes - A Survey Approach”. In: Oct. 2002.

[37] Patrick Herbke, Anish Sapkota, and Sid Lamichhane. “Lifecycle Management of ResumÃ©s with Decentralized Identifiers and Verifiable Credentials”. In: *2024 6th Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS)* (Oct. 2024), pp. 1–3.

[38] Melody A. Hertzog. “Considerations in determining sample size for pilot studies”. In: *Research in Nursing & Health* 31.2 (Jan. 2008), pp. 180–191.

[39] Jeffery Hill et al. “Educator’s blueprint: A how-to guide for collecting validity evidence in survey- based research”. In: *AEM Education and Training* 6.6 (Dec. 2022).

[40] Sunshine Hillygus et al. *Diagnosing survey response quality*. 2022.

[41] Mahmood Hosseini et al. “Four reference models for transparency requirements in information systems”. In: *Requirements Engineering* 23.2 (Mar. 2017), pp. 251–275.

[42] Fadhl Hujainah et al. “Software Requirements Prioritisation: A systematic literature review on significance, stakeholders, techniques and challenges”. In: *IEEE Access* 6 (Jan. 2018), pp. 71497–71523.

[43] Ayei Ibor et al. “Considerations for trustworthy cross-border interoperability of digital identity systems in developing countries”. In: *AI & Society* 40.4 (Aug. 2024), pp. 2729–2750.

[44] Georgy Ishmaev et al. “Value Sensitive Design for Self-Sovereign Identity Solutions: Conceptual investigation of UNLock Use Case”. In: *Deleted Journal* 2.2 (June 2023).

[45] George A. Johanson and Gordon P. Brooks. “Initial Scale development: sample size for pilot studies”. In: *Educational and Psychological Measurement* 70.3 (Dec. 2009), pp. 394–400.

[46] Joachim Karlsson et al. “A Cost-Value approach for prioritizing requirements”. In: *IEEE Software* (1997), pp. 68–69.

[47] Javed Ali Khan et al. “Comparison of requirement prioritization techniques to find best prioritization technique”. In: *International Journal of Modern Education and Computer Science* 7.11 (Nov. 2015), pp. 53–59.

[48] Malcolm Koo and Shih-Wei Yang. “Questionnaire Use and development in health research”. In: *Encyclopedia* 5.2 (May 2025), p. 65.

[49] Jon A. Krosnick and Stanley Presser. *Handbook of Survey Research*. 2nd ed. Apr. 2010.

[50] Gabriella Laatikainen, Mekhail Mustak, and Nicky Hickman. "Self-sovereign identity adoption: Antecedents and potential outcomes". In: *Technology in Society* 82 (Feb. 2025), p. 102859.

[51] Yorick Last and Patricia Arias Cabarcos. "Vision: Towards True User-Centric Design for Digital Identity Wallets". In: *Symposium on Usable Security and Privacy (USEC) 2025* (Jan. 2025).

[52] Jonas Lau and Annie Tran. *UXR point of view on product feature prioritization prior to Multi-Million engineering commitments*. June 2025. URL: http://arxiv.org/abs/2506.15294.

[53] Bettina Laugwitz, Theo Held, and Martin Schrepp. *Construction and evaluation of a user experience questionnaire*. Jan. 2008, pp. 63–76.

[54] Julie Anne Lee, Geoffrey N. Soutar, and Jordan Louviere. "Measuring values using best-worst scaling: The LOV example". In: *Psychology and Marketing* 24.12 (Nov. 2007), pp. 1043–1058.

[55] Jacie L. Lemos et al. "Time-dependent, patient-centered perceptions of quality measures for total joint arthroplasty: a cross-sectional, choice modeling study". In: *BMC Musculoskeletal Disorders* 26.1 (Jan. 2025), p. 41.

[56] James R. Lewis. "Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies". In: *International Journal of Human-Computer Interaction* 14.3-4 (Sept. 2002), pp. 463–488.

[57] None Soo Ling Lim and A. Finkelstein. "StakeRare: using social networks and collaborative filtering for Large-Scale requirements elicitation". In: *IEEE Transactions on Software Engineering* 38.3 (Apr. 2011), pp. 707–735.

[58] Yue Liu et al. "Design patterns for blockchain-based Self-Sovereign Identity". In: *Proceedings of the European Conference on Pattern Languages of Programs 2020* (July 2020).

[59] Anastasios Liveretos and Milena Lazarova. "Decentralized Identity and Verifiable Credentials - Trends and Advancement". In: *2024 32nd National Conference with International Participation (TELECOM)* (Nov. 2024), pp. 1–4.

[60] Torsten Lodderstedt et al. *OpenID for Verifiable Credential Issuance 1.0*. Sept. 2025. URL: https://openid.net/specs/openid-4-verifiable-credential-issuance-1_0.html.

[61] Jordan J. Louviere, Terry N. Flynn, and A. A. J. Marley. *Best-Worst Scaling*. Sept. 2015.

[62] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. "Internet Users' Information Privacy Concerns (IUIPC): the construct, the scale, and a causal model". In: *Information Systems Research* 15.4 (Dec. 2004), pp. 336–355.

[63] Aigul Mavletova. "Data quality in PC and mobile web surveys". In: *Social Science Computer Review* 31.6 (Apr. 2013), pp. 725–743.

[64] Carlo Mazzocca et al. "A survey on decentralized identifiers and verifiable credentials". In: *IEEE Communications Surveys & Tutorials* (Jan. 2025), p. 1.

[65] D. Harrison McKnight, Vivek Choudhury, and Charles Kacmar. "Developing and Validating trust Measures for e-Commerce: An Integrative Typology". In: *Information Systems Research* 13.3 (Sept. 2002), pp. 334–359.

[66] Cameron McPhee et al. *Data Quality Metrics for Online Samples: Considerations for Study Design and Analysis*. Tech. rep. Nov. 2022.

[67] Natalja Menold and Tenko Raykov. "On the Relationship Between Item Stem Formulation and Criterion Validity of Multiple-Component Measuring Instruments". In: *Educational and Psychological Measurement* 82.2 (Feb. 2021), pp. 356–375.

[68] Emmanuel Mkpojiogu and Nor Laily Hashim. "Quality-Based Prioritization: An Approach for Prioritizing Software Requirements". In: *Journal of Communication Electronic and Computer Engineering* 9 (June 2017), pp. 17–21.

[69] Nazila Gol Mohammadi et al. "An Analysis of Software Quality Attributes and Their Contribution to Trustworthiness". In: *Proceedings of the 3rd International Conference on Cloud Computing and Services Science (CloudSecGov-2013), pages 542-552* (Jan. 2013), pp. 542–552.

[70] Beth Morling. *Research methods in Psychology*. 4th ed. 2020.

[71] Axel C. Mühlbacher et al. "Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview". In: *Health Economics Review* 6.1 (Jan. 2016).

[72] Alexander Mühle et al. "A survey on essential components of a self-sovereign identity". In: *Computer Science Review* 30 (Oct. 2018), pp. 80–86.

[73] Julius Olatunji Okesola et al. "Reviewing the Role of Stakeholders in Requirement Engineering: A Stakeholder's Theory perspective". In: *Asian Journal of Scientific Research* 13.1 (Dec. 2019), pp. 1–8.

[74] Kazeem Olanrewaju. "Business Data Breaches-Impact on brand reputation and employee Integrity: A case study of Desjardins in Canada". In: *Texila international journal of academic research* 12.02 (Apr. 2025).

[75] Thomas Olsson, Séverine Sentilles, and Efi Papatheocharous. "A systematic literature review of empirical research on quality requirements". In: *Requirements Engineering* 27.2 (Feb. 2022), pp. 249–271.

[76] Daniela Pöhn, Michael Grabatin, and Wolfgang Hommel. "Analyzing the threats to Blockchain-Based Self-Sovereign identities by conducting a literature survey". In: *Applied Sciences* 14.1 (Dec. 2023), p. 139.

[77] Daniela Pöhn, Michael Grabatin, and Wolfgang Hommel. "eID and Self-Sovereign Identity Usage: An Overview". In: *Electronics* 10.22 (Nov. 2021), p. 2811.

[78] Vikas Prajapati. "Blockchain-Based Decentralized Identity Systems: A Survey of Security, Privacy, and Interoperability". In: *International Journal of Innovative Science and Research Technology (IJISRT)* (Mar. 2025), pp. 1011–1020.

[79] Stanley Presser et al. *Methods for testing and evaluating survey questionnaires*. June 2004.

[80] Carolyn C Preston and Andrew M Colman. "Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences". In: *Acta Psychologica* 104.1 (Mar. 2000), pp. 1–15.

[81] Šeila Bećirović Ramić et al. "Selective disclosure in digital credentials: A review". In: *ICT Express* 10.4 (May 2024), pp. 916–934.

[82] Felipe Ramos et al. "Evaluating software developers' acceptance of a tool for supporting agile Non-Functional requirement elicitation". In: *Proceedings/Proceedings of the ... International Conference on Software Engineering and Knowledge Engineering* 2019 (July 2019), pp. 26–31.

[83] Najia Saher, Fauziah Baharom, and Rohaida Romli. "A Review of Requirement Prioritization Techniques in Agile Software Development". In: *Knowledge Management International Conference (KMICe) 2018* (2018), http://www.kmice.cms.net.my.

[84] Abylay Satybaldy. *Usability Evaluation of SSI digital wallets*. Jan. 2023, pp. 101–117.

[85] Riccardo Scarpa et al. "Exploring Scale Effects of Best/Worst Rank Ordered Choice Data to Estimate Benefits of Tourism in Alpine Grazing Commons". In: *American Journal of Agricultural Economics* 93.3 (2011), pp. 813–828.

[86] Frederico Schardong and Ricardo Custódio. "Self-Sovereign Identity: A Systematic Review, Mapping and Taxonomy". In: *Sensors* 22.15 (July 2022), p. 5641.

[87] Martin Schrepp. *On the Usage of Cronbach's Alpha to Measure Reliability of UX Scales - JUX*. Aug. 2020. URL: https://uxpajournal.org/cronbachs-alpha-reliability-ux-scales/#:~:text=As%20with%20any%20other%20statistic,0.5%20%28Unacceptable.

[88] Daria Schumm, Katharina O. E. Müller, and Burkhard Stiller. *Are we there yet? A study of decentralized identity applications*. Mar. 2025. URL: http://arxiv.org/abs/2503.15964.

[89] Anne L.R. Schuster et al. "The rise of best-worst scaling for prioritization: A transdisciplinary literature review". In: *Journal of Choice Modelling* 50 (Jan. 2024), p. 100466.

[90] Johannes Sedlmeir et al. "Digital identities and verifiable credentials". In: *Business & Information Systems Engineering* 63.5 (Oct. 2021), pp. 603–613.

[91] Ahmed Seffah et al. "Usability measurement and metrics: A consolidated model". In: *Software Quality Journal* 14.2 (May 2006), pp. 159–178.

[92] H. Jeff Smith, Sandra J. Milberg, and Sandra J. Burke. "Information Privacy: Measuring Individuals' Concerns about Organizational Practices". In: *MIS Quarterly* 20.2 (June 1996), p. 167.

[93] Reza Soltani, Uyen Trang Nguyen, and Aijun An. "A Survey of Self-Sovereign Identity Ecosystem". In: *Security and Communication Networks* 2021 (July 2021), pp. 1–26.

[94] Gail M. Sullivan and Anthony R. Artino. "Analyzing and interpreting data from Likert-Type scales". In: *Journal of Graduate Medical Education* 5.4 (Dec. 2013), pp. 541–542.

[95] Hamed Taherdoost. "Designing a questionnaire for a research paper: A comprehensive guide to design and develop an effective questionnaire". In: *Asian Journal of Managerial Science* 11.1 (Apr. 2022), pp. 8–16.

[96] Oliver Terbu et al. *OpenID for Verifiable Presentations 1.0*. July 2025. URL: https://openid.net/specs/openid-4-verifiable-presentations-1_0-final.html.

[97] Špela Čučko, Vid Keršič, and Muhamed Turkanović. "Towards a catalogue of Self-Sovereign identity design patterns". In: *Applied Sciences* 13.9 (Apr. 2023), p. 5395.

[98] Špela Čučko et al. "Towards the classification of Self-Sovereign Identity properties". In: *IEEE Access* 10 (Jan. 2022), pp. 88306–88329.

[99] W3. *Bitstring Status List v1.0*. May 2025. URL: https://www.w3.org/TR/vc-bitstring-status-list/.

[100] W3. *Verifiable Credentials Data Model v2.0*. May 2025. URL: https://www.w3.org/TR/vc-data-model-2.0/.

[101] Colin Werner et al. *Continuous Non-Functional requirements: practices, opportunities, and trade-offs for small, agile organizations*. 2021. URL: https://zenodo.org/record/3376343.

[102] Gordon Willis. *Cognitive interviewing as a tool for improving consent and communication*. Jan. 2006. URL: https://doi.org/10.1037/e538062007-001.

[103] Sandhya Yaddanapudi and Ln Yaddanapudi. "How to design a questionnaire". In: *Indian Journal of Anaesthesia* 63.5 (Jan. 2019), p. 335.

[104] Wei Yao et al. "Establishing a baseline for evaluating Blockchain-Based Self-Sovereign Identity Systems: a systematic approach to assess capability, compatibility and interoperability". In: *Proceedings of the 2024 6th Blockchain and Internet of Things Conference* (July 2024), pp. 108–119.

[105] Eric Yu and Luiz Cysneiros. "Designing for Privacy and Other Competing Requirements". In: (Nov. 2002).

[106] Razieh Nokhbeh Zaeem et al. "Blockchain-Based Self-Sovereign Identity: Survey, Requirements, Use-Cases, and Comparative Study". In: *IEEE/WIC/ACM International Conference on Web Intelligence* (Dec. 2021), pp. 128–135.

[107] University of Zurich. *Data protection in research projects*. URL: https://www.rud.uzh.ch/en/angebot/datenschutzrecht/research.html.

[108] University of Zurich. *Friedman-Test*. URL: https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/unterschiede/zentral/friedman.html.

[109] University of Zurich. *Kruskal-Wallis-Test*. URL: https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/unterschiede/zentral/kruskal.html.

[110] University of Zurich. *Legal and ethical guidelines*. URL: https://www.openscience.uzh.ch/en/research-integrity/legal-and-ethical-guidelines.html.

[111] University of Zurich. *Methodenberatung*. URL: https://www.methodenberatung.uzh.ch/de/datenanalyse_spss.html.

[112] University of Zurich. *Pearson Chi-Quadrat-Test (Kontingenzanalyse)*. URL: https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/zusammenhaenge/pearsonzush.html.

[113] University of Zurich. *Rangkorrelation nach Spearman*. URL: https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/zusammenhaenge/rangkorrelation.html.

[114] University of Zurich. *Research involving Human Beings*. URL: https://www.research.uzh.ch/en/procedures/research-on-humans.html.

[115] University of Zurich. *Wilcoxon-Test*. URL: https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/unterschiede/zentral/wilkoxon.html.

[116] Guy Zyskind, Oz Nathan, and Alex 'Sandy' Pentland. "Decentralizing Privacy: Using Blockchain to Protect Personal Data". In: *2015 IEEE Security and Privacy Workshops* (May 2015), pp. 180–184.

Abbreviations

AHP	Analytic Hierarchy Process
BWS	Best-Worst Scaling
CFIP	Concern for Information Privacy
DID	Decentralized Identifier
DI	Decentralized Identity
FI	Functional Importance
FR	Functional Requirement
IdP	Identity Provider
IUIPC	Internet Users' Information Privacy Concerns
MNL	Multinomial Logit Model
NFR	Non-Functional Requirement
PI	Problem Importance
PSSUQ	Post-Study System Usability Questionnaire
QUIM	Quality in Use Integrated Measurement
RE	Requirements Engineering
RQ	Research Question
SD	Standard Deviation
SeBIS	Security Behavior Intentions Scale
SQRI	Software Quality Requirements Importance
SSI	Self-Sovereign Identity
SUS	System Usability Scale
UEQ	User Experience Questionnaire
UMUX	Usability Metric for User Experience
VC	Verifiable Credential
W3C	World Wide Web Consortium
W	Wilcoxon
TLS	Transport Layer Security

List of Figures

List of Tables

Listings

Appendix A

Contents of the NFR Mapping & Questionnaire Design

A.1 NFR Categorization

Key	Quality	Data Owner	Data Owner Responsibility	Justification	Verifier	Verifier Responsibility	Justification	Issuer	Issuer Responsibility	Justification	Reasoning for Responsibility Level Assignment	Ownership
NFR1	Accessibility	☑	PRIMARY	Must be able to access and retrieve data about their own credential or DID. It is paramount that the holder has continuous access to the identity data in the wallet, and can obtain identity attributes instantly from issuers (Čučko et al., 2022).	☑	TERTIARY	Must be able to access and retrieve data about credential issuer (Soltani et al., 2021).	☐		The issuer needs dependable access to manage credential schemas effectively, which are essential for defining the structure and semantics of the credentials they issue, ensuring that verifiable data registries are up to date (Soltani et al., 2021).	The Data Owner directly accesses their data. The Verifier and Issuer access data about others (issuers, schemas) to facilitate their main functions.	Data Owner
NFR3	Autonomy	☑	PRIMARY	Must be able to manage identity data independently of a third party. The data owner must possess the autonomous capability to create, store, update, and share their identifiers and credentials independently, without requiring permission from a central authority; this "self-managed identity" ability is the essence of autonomy (Čučko et al., 2022)	☐			☐			Autonomy is a core, defining responsibility of the Data Owner.	Data Owner
NFR4	Availability	☑	PRIMARY	Identity data must be available at all time. It is paramount that the holder has continuous access to the identity data in the wallet, and can obtain identity attributes instantly from issuers. Storing identity data by users themselves largely preserves this property (Čučko et al., 2022).	☐			☐			Availability of identity data is a core responsibility resting on the Data Owner's ability to access their wallet and credentials.	Data Owner
NFR6	Consent	☑	PRIMARY	Entities should be able to give deliberate and well-understood consent for the usage/sharing of their identity data, and should be able to withdraw/revoke that consent at a later date (Čučko et al., 2022). The data holder's consent is required for every access to their identity data. (Zaeem et al., 2021)	☑	PRIMARY	Must ensure compliance with GDPR. Entities in this role must obtain consent from identity holders for activities such as data collection, processing, storage, etc. Minimal data disclosure forces them to obtain only the minimum amount of data relevant for a single transaction (Čučko et al., 2022)	☑	PRIMARY	Must ensure compliance with GDPR. In accordance with informed consent principles, issuers are obligated to secure explicit consent from subjects before issuing credentials pertaining to their personal information (Čučko et al., 2022).	All three roles have a primary and direct responsibility: the Data Owner to give consent, and the Verifier and Issuer to obtain it for their respective processes.	Data Owner Verifier
NFR7	Control	☑	PRIMARY	Must be able to control access to the identity data. Data owners keep their DIDs, verifiable credentials, and private keys entirely within their own wallet, making it the single source of truth for their identity and letting them decide what to reveal or revoke via verifiable presentations (Čučko et al., 2022). This self-managed model shifts control—though not necessarily legal ownership—from organisations to individuals, aligns with GDPR's user-centric consent principles, and enhances privacy and security (Liveretos & Lazarova, 2024). Users must maintain authority over access to their identity data. (Zaeem et al., 2021).	☐		The verifier has no control over the identity data and hence should not have the possibility to control it in any way (Čučko et al., 2022).	☐			Control is the central, primary responsibility of the Data Owner in SSI.	Data Owner
NFR12	Persistence	☑	PRIMARY	Identity data must remain valid for as long as necessary. It is also significant that the holder's identity is available as long as he/she requires it (Čučko et al., 2022). Identifiers should be persistent and exist for at least as long as it is required by their owner. The availability of identity data persists for a duration determined by the holder, ceasing only upon their explicit removal (Zaeem et al., 2021).	☐			☐			Persistence is a primary responsibility of the Data Owner, who determines how long their identity is needed.	Data Owner
NFR14	Privacy	☑	PRIMARY	Must be able to minimize information required to share and not disclose unnacessary data. Data Owner can preserve their privacy by exposing only the minimum data set required for successful completion of a particular interaction (Čučko et al., 2022).	☑	PRIMARY	Must request minimal information required. Minimal data disclosure is also essential from the perspective of GDPR compliance. Verifiers only acquire and process data required for provisioning a particular service (Čučko et al., 2022).	☑	SECONDARY	Issuers are responsible for storing and processing only the data necessary for issuing credentials, whereas verifiers are limited to acquiring and processing data required for delivering a specific service (Čučko et al., 2022).	Privacy is a shared Primary responsibility for the Data Owner and the Verifier. The principle is only fulfilled when the Verifier requests the minimum data necessary AND the Data Owner selectively discloses it. Both are direct, essential actions. The Issuer has a Secondary role, supporting system-wide privacy by adhering to data minimization principles during its own issuance process.	Data Owner Verifier
NFR15	Protection	☑	PRIMARY	The data owner must ensure that identity data is safeguarded from misuse during storage and transmission. Trust in the technology and a sense of overall protection are essential when handling identity data, ensuring the rights of entities are protected (Čučko et al., 2022). The information must be secure (Zaeem et al., 2021).	☑	SECONDARY	Identity data must be protected against misuse during verification process. Trust in the security of the technology is also critical for the verifier. Identity data must be protected against misuse during verification process (Čučko et al., 2022).	☑	SECONDARY	Identity data must be protected against misuse during issuance of credential. Trust in the security of the technology is also critical for the issuer. Identity data must be protected against misuse during issuance of credential (Čučko et al., 2022).	The Data Owner is primarily responsible for protecting their own wallet/data. The Verifier and Issuer have a secondary responsibility to protect data during their respective processes (verification and issuance).	Data Owner
NFR16	Recoverability	☑	PRIMARY	Must be able to recover their digital identity successfully without having to reacquire previously obtained credentials, in case of loss of keys, phone loss, or digital wallet vulnerability (Čučko et al., 2022). Data owners must be able to recover their digital identities, including keys and credentials, efficiently and securely (Zaeem et al., 2021).	☐			☐			Recoverability is a core, primary responsibility for the Data Owner.	Data Owner System
NFR17	Representation	☑	PRIMARY	As a holder, the ability to create many identities for digital interactions is crucial, since it allows them to interact and present themselves differently (Čučko et al., 2022).	☐			☐			The ability to create different representations for different contexts is a primary function of the Data Owner.	Data Owner
NFR19	Single Source	☑	PRIMARY	Data owners must maintain a single, authoritative source of truth regarding their identity to ensure control and prevent unauthorized data transmission (Čučko et al., 2022).	☐			☐			Maintaining a single source of truth is a core, primary responsibility of the Data Owner.	Data Owner
NFR21	Transparency	☑	PRIMARY	The data owner should possess a comprehensive understanding of all established connections, acquired credentials, data sharing activities, and interaction history (Čučko et al., 2022). According to Zaeem et al. (2021), Data owners should possess a comprehensive understanding of the entities holding their data.	☑	SECONDARY	Information about identity data use must be readily available. Verifiers should operate under transparent systems and algorithms, ensuring that the processes by which they validate credentials and derive trust are clear and understandable to all participants in the ecosystem (Soltani et al., 2021).	☑	SECONDARY	Information about identity data use must be readily available. Issuers must implement transparent systems and algorithms, ensuring that the processes by which they issue credentials are auditable and understandable by relevant stakeholders, including data owners and relying parties (Soltani et al., 2021). Information about identity data use must be readily available.	The Data Owner has a primary need for transparency into how their data is used. The Verifier and Issuer have a secondary responsibility to make their processes transparent to facilitate trust.	Data Owner
NFR10	Existence	☑	SECONDARY	According to Čučko et al., 2022, entities must have an independent existence and should be able to create as many identities as required without the intervention of a third party (Čučko et al., 2022). Crucial ability to create as many identities as required for digital interactions, since it allows an individual to interact and present him/herself in different contexts. (Čučko et al., 2022). DIDs empower entities... with Existence and Representation. Allows him/her to create an identity, i.e., identifier and self-attested attributes, without the intervention of any intermediary. The public permissionless blockchain allows entities to create and register as many identifiers as required (Existence and Representation). Data remains visible to identity holders unless they explicitly request its removal (Zaeem et al., 2021).	☐			☐			The Data Owner has a Secondary responsibility, as their ability to exist and create identities facilitates their primary duties like exercising control. The Issuer is not assigned a responsibility level because NFR10 is defined from a user-centric perspective (creating multiple identities for different contexts). This does not align with the Issuer's need for a single, stable identity to ensure trust, making the NFR inapplicable to their role.	System
NFR2	Authenticity	☑	TERTIARY	The user must carry credentials whose signatures prove they come from a legitimate source and have not been altered. This lets the holder demonstrate, without an online check, that "the digital identities are controlled by their owners and haven't been tampered with" (Čučko et al., 2022)	☑	SECONDARY	Source of identity data must be trustworthy and provable. Authenticity is not a quality the verifier needs to possess; it is the property the verifier checks. As Čučko et al. (2022) explain, authenticity relates to the holder's ability to prove identity, "not from the verification context, which is in the domain of the verifiers." (Čučko et al., 2022)	☑	PRIMARY	Authenticity begins with the issuer. By attaching its own digital signature to every credential, the issuer guarantees the data's origin and integrity. If the issuer's signing process were weak, no holder could later prove authenticity and no verifier could trust the result (Čučko et al., 2022).	The Issuer holds the Primary responsibility as they are the source of authenticity, creating and guaranteeing it by digitally signing the credential. The Data Owner and Verifier play crucial, Secondary roles: the Data Owner facilitates the process by presenting the credential, and the Verifier facilitates trust by checking it.	Issuer
NFR5	Compatibility	☐	TERTIARY	Compatibility with legacy systems "is not that evident to the identity holders," because the wallet abstracts those protocol details away; the user mainly benefits indirectly when credentials simply work (Čučko et al., 2022).	☑	SECONDARY	Identity data must be compatible with legacy system (Čučko et al., 2022). Verifiers need to import and validate credentials issued under a variety of standards and transport protocols (Čučko et al., 2022).	☑	PRIMARY	Identity data must be compatible with legacy system (Čučko et al., 2022). The issuer must create credentials that work with both new SSI wallets and older PKI-based infrastructures, since "identity should be backward compatible with legacy identity systems to ensure quicker acceptance" (Čučko et al., 2022).	The Issuer has the primary responsibility to create compatible credentials. The Verifier has a secondary role in needing to support various standards. The Data Owner is a tertiary beneficiary of this compatibility.	Issuer
NFR8	Cost	☑	TERTIARY	Must have minimal costs for storing, generating proofs, and presentations. The cost of participating in the SSI ecosystem also affects the identity holders' decision to engage and use the technology greatly, whereby, from his/her point of view, it includes mainly digital wallet and transaction costs, as well as learning and required effort (Čučko et al., 2022). The minimization of financial costs is an essential factor for identity owners, issuers, and verifiers (Zaeem et al., 2021).	☑	TERTIARY	Must have minimal costs for verification. Additional implementation or integration costs must be considered (Goel & Rahulamathavan, 2024). The minimization of financial costs is an essential factor for identity owners, issuers, and verifiers (Zaeem et al., 2021).	☑	TERTIARY	Must have minimal costs for issuing verifiable credentials. Similar to verifiers, issuers face implementation costs when integrating SSI into existing systems, requiring a balance between security, compliance, and economic feasibility (Čučko et al., 2022; Goel & Rahulamathavan, 2024). The minimization of financial costs is an essential factor for identity owners, issuers, and verifiers (Zaeem et al., 2021).	Cost is a Tertiary responsibility for all three roles. While it is a primary consideration for each actor, none are directly responsible for creating or guaranteeing a cost-effective system. Instead, cost is determined by the system's design and technology choices. Therefore, all actors are indirect beneficiaries who rely on the system's design to be affordable.	System

Key	Quality	Data Owner	Data Owner Responsibility	Justification	Verifier	Verifier Responsibility	Justification	Issuer	Issuer Responsibility	Justification	Reasoning for Responsibility Level Assignment	Ownership
NFR9	Decentralization	☑	TERTIARY	Must be able to generate multiple DIDs independently, and be able to resolve them without a central authority. Should not rely on centralized elements for storage. Storing identity data by users themselves largely preserves this property. The core concept revolves around the data owner remaining in full control of their identifier and associated verification material (Garzon et al., 2024). This ensures that no entity other than the data owner is involved in the creation and management of their DID (Garzon et al., 2024).	☑	TERTIARY	Should not rely on centralized elements for verification.Verifiers should not depend on any centralized infrastructure: by resolving the issuer's DID document in a distributed registry and checking the signature locally, they can authenticate claims directly while fully respecting the holder's autonomy and control (Garzon et al., 2024; W3C, 2023).	☑	TERTIARY	Should not rely on centralized elements for issuance. Issuers publish their own DID and issue / revoke credentials via decentralised registries, so issuance is not tied to a central identity provider (Liveretos & Lazarova, 2024; Soltani et al., 2021)	Decentralization is a Tertiary responsibility for all three roles. Similar to 'Cost,' it is a fundamental property of the system's architecture, not a direct responsibility of the actors. The actors don't create the decentralized environment, but they all benefit from it, as they rely on the system's design to grant them autonomy and remove single points of failure.	System
NFR11	Interoperability	☑	TERTIARY	Interoperability ensures that the Data Owner's identity is not locked into a single system or vendor (Liveretos & Lazarova, 2024). They should be able to seamlessly interact with various services and organizations that accept Verifiable Credentials (Čučko et al., 2022). The data owner requires the ability to access a wide range of public and private services, ensuring compatibility across various programming languages, blockchains, vendors, platforms, networks, legal jurisdictions, geographies, cryptographic methods, and hardware, as well as over extended periods (Zaeem et al., 2021).	☑	SECONDARY	Must be to verify credentials from different platforms and services. Interoperability is crucial for Verifiers to avoid being limited to accepting credentials from a closed ecosystem (Liveretos & Lazarova, 2024). A Verifier's ability to work across different systems builds trust that the Verifier will not become obsolete.	☑	PRIMARY	Must be able to issue credentials that is usable across different platforms and services (Liveretos & Lazarova, 2024).	The Issuer has the primary duty to issue interoperable credentials. The Verifier has a secondary duty to accept them. The Data Owner is the tertiary beneficiary of a seamless experience.	Issuer
NFR13	Portability	☑	TERTIARY	Identity data should be easily transportable from one location to another (Soltani et al., 2021). The facility to port one's digital identity credentials across diverse platforms and systems is essential (Zaeem et al., 2021). Must be able to move their identity data. He/she can transfer identity data from one device or wallet to another. Transferring identity data from one wallet or device to another enables Portability (Čučko et al., 2022).	☐			☐			Portability is a Tertiary responsibility for the Data Owner. Although the owner performs the action of moving their data, their ability to do so depends on the system's architecture and standards. They rely on wallet developers and the ecosystem to enable portability, and thus are the indirect beneficiaries of this system property.	System
NFR18	Security	☑	TERTIARY	From a security standpoint, it is essential for data owners to have confidence in the technology and the parties involved, as well as an overall sense of security when transmitting and storing identity data (Čučko et al., 2022). Data owners must also feel secure, trusting that data breaches, misuse, errors, and other security threats are effectively managed.	☑	TERTIARY	Identity data must be secure during verification. Trust in the security of the technology is also critical for the verifier. Authentication is a critical step for both roles, to prevent misuse and minimize security threats. Verifiers are required to authenticate and authorize identity holders to ensure they are granted appropriate access to services (Čučko et al., 2022). Verifiers need the ability to identify the issuer of a credential, confirm its integrity, and verify that it remains valid, unexpired, and unrevoked (Čučko et al., 2022).	☑	TERTIARY	Identity data must be secure during issuance of credential. Trust in the security of the technology is also critical for the issuer. Authentication is a critical step for both roles, to prevent misuse and minimize security threats. Identity data must be secure during issuance of credential. To ensure the integrity of verifiable credentials, issuers are obligated to authenticate identity holders before affirming identity attributes (Čučko et al., 2022).	All three roles have a primary responsibility for security within their own domain: the owner securing their wallet, the issuer securing the issuance process, and the verifier securing the verification process. Security is ultimately a Primary responsibility for all actors. While they have a Tertiary reliance on the system's secure design (e.g., strong cryptography), this is insufficient. Security is only achieved through the Primary, direct actions of each actor: the Data Owner must protect their keys, the Issuer must properly vet identities, and the Verifier must perform correct authentication. A failure in these direct duties by any actor is critical, making their role primary.	Data Owner Issuer Verifier System
NFR20	Standard	☐	TERTIARY	Standard is not that evident to the identity holders (Čučko et al., 2022)	☑	SECONDARY	The verifier may decide that certain requests can only be answered if the credential has certain schemas and/or is issued by certain issuers (Bochnia et al., 2024). In this case standards are relevant.	☑	PRIMARY	Issuers possess the capability to manage different versions of schemas (Bochnia et al., 2024). The issuer also has the responsibility of issuing credentials and schemas in compliance with accepted standards (Liveretos & Lazarova, 2024).	The Issuer has the primary responsibility to comply with standards. The Verifier has a secondary role in using these standards. The Data Owner is a tertiary beneficiary.	Issuer
NFR22	Usability	☑	TERTIARY	Must be able to use the identity data efficiently and intuitively (Mazzocca et al., 2025). Usability of decentralized identity systems is critical for mass adoption, necessitating solutions that simplify user interactions and reduce the complexity associated with managing digital identities. It is crucial that the wallet not only supports the operations mentioned but also offers ease of use (Čučko et al., 2022).	☐			☐			Usability is a Tertiary responsibility for all actors. The Data Owner is the main beneficiary of a usable system but relies on designers and developers to implement it. Issuers and Verifiers also indirectly benefit from the Data Owner's good user experience, as it promotes wider adoption and smoother interactions.	System
NFR23	User Experience	☑	TERTIARY	Identity management process must be simple, consistent, and user-friendly and should offer good user experience (Čučko et al., 2022)	☐			☐			User Experience is a Tertiary responsibility for the Data Owner. As with Usability, this is a quality that the user benefits from but does not control. They rely entirely on the system's designers to ensure a positive experience.	System
NFR24	Verifiability	☐	TERTIARY	No specific relevance. Not assigned for this property in Čučko et al.'s taxonomy (Čučko et al., 2022). While holders benefit from using verifiable artefacts, the responsibility for making credentials verifiable (issuer) and for performing the verification (verifier) lies with the other two roles.	☑	PRIMARY	Must be able to verify data. The system shall provide the organization as a verifier with the ability to request credentials from the holder (Bochnia et al., 2024). Verifiers must be able to independently verify the validity and authenticity of credentials presented to them, ensuring that the claims made are trustworthy and have not been tampered with (Mazzocca et al., 2025).	☑	SECONDARY	The data must be verifiable. The issuer must produce credentials that remain verifiable throughout their lifetime by digitally signing each credential, publishing the corresponding public key (for example, in a DID document), and updating status or revocation lists when necessary so verifiers can confirm that the credential is still valid (Čučko et al., 2022).	The Verifier holds the Primary responsibility, as they perform the core action of checking the credential. The Issuer plays a crucial Secondary role by facilitating this verification; their duty is to produce a verifiable artifact for the Verifier to check. The Data Owner is the Tertiary beneficiary of this process.	Verifier

A.2 NFR Operationalization & Overview of Survey Questions

Table A.1: Operationalized NFR items for Identity Holders (Users)

NFR	Role	Item Type	Design Pattern	Survey Item
Accessibility	Holder	Functional Importance	DID Management Patterns - ”Identifier Registry” entry, Table 1, [58]	I need to access my identity data whenever I want.
Accessibility	Holder	Problem Importance	DID Management Patterns - ”Identifier Registry” entry, Table 1, [58]	Limited access to my identity data when I need it would be acceptable to me.
Autonomy	Holder	Functional Importance	Key Management Patterns - ”Master and Sub Key Generation”, [58] Decentralised Identifiers and Cryptographic Keys category - ”DID Controller”, [97]	I want to manage my digital identity myself, without relying on a central authority.
Autonomy	Holder	Problem Importance	Key Management Patterns - ”Master and Sub Key Generation”, [58] Decentralised Identifiers and Cryptographic Keys category - ”DID Controller”, [97]	If I had to ask someone else to change or share my digital identity, it would be fine for me.
Availability	Holder	Functional Importance	DID Management Patterns - ”Dual Resolution”, [58] Key Management Patterns - ”Hot and Cold Wallet Storage”, [58] Wallet and Off-chain Storage group, Pattern ”Remote Storage”, [97]	My digital identity should be available whenever I need it, even if some services are down.
Availability	Holder	Problem Importance	DID Management Patterns - ”Dual Resolution”, [58] Key Management Patterns - ”Hot and Cold Wallet Storage”, [58] Wallet and Off-chain Storage group, Pattern ”Remote Storage”, [97]	Being prevented from using my credentials due to a service outage would be acceptable for me.
Consent	Holder	Functional Importance	Verifiable Credentials and Presentations category - discussion of ”time-constrained access”, ”one-off access”, [97] Credential Design Patterns - ”Time-Constrained Access”	I want services to ask for my permission each time they use my identity information.
Consent	Holder	Problem Importance	Verifiable Credentials and Presentations category - discussion of ”time-constrained access”, ”one-off access”, [97] Credential Design Patterns - ”Time-Constrained Access”	If my identity information were used without my permission, it would be manageable for me.
Control	Holder	Functional Importance	Credential Design Patterns - ”Selective Content Generation”, [58] Verifiable Credentials and Presentations category - ”Selective Content Generation”, ”Selective Content Disclosure”, [97]	I want to choose exactly which parts of my digital identity I show or hide.
Control	Holder	Problem Importance	Credential Design Patterns - ”Selective Content Generation”, [58] Verifiable Credentials and Presentations category - ”Selective Content Generation”, ”Selective Content Disclosure”, [97]	If I lacked control over which identity details were shared, it would be acceptable to me.
Persistence	Holder	Functional Importance	Wallet and Off-Chain Storage, [97]	I want my digital-identity credentials to stay valid and usable for as long as I need them.
Persistence	Holder	Problem Importance	Wallet and Off-Chain Storage, [97]	If my credentials expired or disappeared before I chose to remove them, that would be acceptable for me.
Privacy	Holder	Functional Importance	Credential Design Patterns - ”Selective Content Generation”, [58] Verifiable Credentials and Presentations - ”Selective Content Generation”, ”Selective Content Disclosure”, [97]	I want to share only the minimum details about myself when I prove my identity.

NFR	Role	Item Type	Design Pattern	Survey Item
Privacy	Holder	Problem Importance	Credential Design Patterns - "Selective Content Generation", [58] Verifiable Credentials and Presentations - "Selective Content Generation", "Selective Content Disclosure", [97]	If I had to reveal more personal information than necessary, I would be fine.
Protection	Holder	Functional Importance	Key Management Patterns - "Hot and Cold Wallet Storage", [58] Wallet and Off-Chain Storage group - "Hot and Cold Wallet Storage", [97]	I need my identity data to be protected from unauthorized access or tampering during storage and transmission.
Protection	Holder	Problem Importance	Key Management Patterns - "Hot and Cold Wallet Storage", [58] Wallet and Off-Chain Storage group - "Hot and Cold Wallet Storage", [97]	If someone could read or change my identity information without permission, I would accept that.
Recoverability	Holder	Functional Importance	Decentralised Identifiers and Cryptographic Keys category, "Delegate List", [97] Key Management Patterns, "Key Shards", [58]	I want to be able to restore my digital identity if I lose my phone or keys.
Recoverability	Holder	Problem Importance	Decentralised Identifiers and Cryptographic Keys category, "Delegate List", [97] Key Management Patterns, "Key Shards", [58]	If I were unable to recover my identification after losing my device or keys, it would be manageable for me.
Representation	Holder	Functional Importance	Decentralised Identifiers and Cryptographic Keys category, "Multiple Registration", [58] DID Management Patterns, "Multiple Registration", [97]	I want separate digital identities for different situations.
Representation	Holder	Problem Importance	Decentralised Identifiers and Cryptographic Keys category, "Multiple Registration", [58] DID Management Patterns, "Multiple Registration", [97]	I would be comfortable using a single digital identity in every situation.
Single Source	Holder	Functional Importance	Trusted Registries category, "Blockchain Anchor", [97] DID Management Patterns, "Identifier Registry", [58]	I want to have one reliable place that holds the official version of my identity.
Single Source	Holder	Problem Importance	Trusted Registries category, "Blockchain Anchor", [97] DID Management Patterns, "Identifier Registry", [58]	If different services had conflicting versions of my identity, I would be okay with that.
Transparency	Holder	Functional Importance	Trusted Registries category, "Blockchain Anchor", [97] Credential Design Patterns, "Blockchain Anchor", [58]	I want to see who has accessed or shared my identity information and when it happened.
Transparency	Holder	Problem Importance	Trusted Registries category, "Blockchain Anchor", [97] Credential Design Patterns, "Blockchain Anchor", [58]	I could tolerate the uncertainty of which services have utilized my identifying information.
Existence	Holder	Functional Importance	DID Management Patterns, "Multiple Registration", [58] Decentralised Identifiers and Cryptographic Key, "Multiple Registration", [97]	I want to be able to create a new digital identity whenever I choose, without asking anyone.
Existence	Holder	Problem Importance	DID Management Patterns, "Multiple Registration", [58] Decentralised Identifiers and Cryptographic Key, "Multiple Registration", [97]	It would be fine for me if I needed approval from a third party before creating a new digital identity.
Authenticity	Holder	Functional Importance	Trusted Registries category, "Blockchain Anchor", [97] DID Management Patterns, "Identifier Registry", [58]	I want my credentials to carry a signature that proves they are genuine and untouched.
Authenticity	Holder	Problem Importance	Trusted Registries category, "Blockchain Anchor", [97] DID Management Patterns, "Identifier Registry", [58]	If I had to use credentials that might be fake or altered, I could cope with it.
Compatibility	Holder	Functional Importance	DID Management Patterns, "Dual Resolution", [58]	I like it when my digital credentials work smoothly across different websites and apps.
Compatibility	Holder	Problem Importance	DID Management Patterns, "Dual Resolution", [58]	I could handle my digital credentials working in some places but failing in others.

NFR	Role	Item Type	Design Pattern	Survey Item
Cost	Holder	Functional Importance	Key Management Patterns, "Hot and Cold Wallet Storage", [58] Wallet and Off-Chain Storage group, "Remote Storage", [97]	I prefer identity solutions that let me store and use my credentials without extra fees.
Cost	Holder	Problem Importance	Key Management Patterns, "Hot and Cold Wallet Storage", [58] Wallet and Off-Chain Storage group, "Remote Storage", [97]	I could tolerate paying high fees to keep or present my credentials.
Decentralization	Holder	Functional Importance	Trusted Registries group, "Blockchain Anchor", [97] DID Management Patterns, "Identifier Registry", [58]	I want to create and use new digital identities without depending on any single company or server.
Decentralization	Holder	Problem Importance	Trusted Registries group, "Blockchain Anchor", [97] DID Management Patterns, "Identifier Registry", [58]	I would be fine relying on one provider to store or verify my identity.
Interoperability	Holder	Functional Importance	DID Management Patterns, "Blockchain and Social Media Account Pair", [58]	I want my digital credentials to work on any website or app, no matter which platform or provider the website or app uses.
Interoperability	Holder	Problem Importance	DID Management Patterns, "Blockchain and Social Media Account Pair", [58]	Having credentials that work in one place but fail on other services would be acceptable to me.
Portability	Holder	Functional Importance	Wallet and Off-Chain Storage category, "Remote Storage", [97]	I want to move my digital credentials to another phone or wallet whenever I switch devices.
Portability	Holder	Problem Importance	Wallet and Off-Chain Storage category, "Remote Storage", [97]	It would be manageable for me if my credentials failed to transfer to a new device.
Security	Holder	Functional Importance	Key Management Patterns, "Key Shards", [58]	I need assurance that both the technology and its operators secure my identity data against threats.
Security	Holder	Problem Importance	Key Management Patterns, "Key Shards", [58]	I could handle the possibility of my identity data being exposed in a breach.
Standard	Holder	Functional Importance	Trusted Registries category, "Trusted Schemas Registry", [97]	I prefer credentials that follow common standards so every service can read them.
Standard	Holder	Problem Importance	Trusted Registries category, "Trusted Schemas Registry", [97]	I would accept credentials that violate standards and fail various services.
Usability	Holder	Functional Importance	Key Management Patterns, "Hot and Cold Wallet Storage", [58]	Using my digital identity should be quick and straightforward.
Usability	Holder	Problem Importance	Key Management Patterns, "Hot and Cold Wallet Storage", [58]	Having to go through many steps to use my digital identity would be acceptable to me.
User Experience	Holder	Functional Importance		I want managing my digital identity to feel simple and intuitive.
User Experience	Holder	Problem Importance		I would find it acceptable to use a confusing or complicated identity management app.
Verifiability	Holder	Functional Importance	Trusted Registries category, "Status Registry", [97]	I want the people or services I interact with to check my digital credentials instantly, without asking me for extra proof.
Verifiability	Holder	Problem Importance	Trusted Registries category, "Status Registry", [97]	A delay in the verification of my credentials would be tolerable.

Table A.2: Operationalized NFR items for Verifiers

NFR	Role	Item Type	Design Pattern	Survey Item
Consent	Verifier	Functional Importance	Verifiable Credentials and Presentations category, "time-constrained access", "one-off access", [97] Credential Design Patterns, "Time-Constrained Access", [58]	Before using someone's identity data, I want to receive clear consent from the holder every time.
Consent	Verifier	Problem Importance	Verifiable Credentials and Presentations category, "time-constrained access", "one-off access", [97] Credential Design Patterns, "Time-Constrained Access", [58]	Processing identity data without explicit consent from the holder would be fine for my organisation.
Privacy	Verifier	Functional Importance	Credential Design Patterns, "Selective Content Generation", [58] Verifiable Credentials and Presentations category, "Selective Content Generation", "Selective Content Disclosure", [97]	When I check someone's identity, I want to see only the data that are strictly needed for my service.
Privacy	Verifier	Problem Importance	Credential Design Patterns, "Selective Content Generation", [58] Verifiable Credentials and Presentations category, "Selective Content Generation", "Selective Content Disclosure", [97]	Requesting more personal data than necessary would be acceptable for our service.
Verifiability	Verifier	Functional Importance	Trusted Registries category, "Status Registry", [97]	I need to confirm on my own that any credential I receive is genuine and still valid.
Verifiability	Verifier	Problem Importance	Trusted Registries category, "Status Registry", [97]	I trust credentials regardless of whether I can verify their validity myself.
Authenticity	Verifier	Functional Importance	Trusted Registries group, "Blockchain Anchor", [97] Credential Design Patterns, "Blockchain Anchor", [58]	I must be able to prove that every credential I check is genuine and unaltered.
Authenticity	Verifier	Problem Importance	Trusted Registries group, "Blockchain Anchor", [97] Credential Design Patterns, "Blockchain Anchor", [58]	I could accept credentials without verifying their authenticity.
Compatibility	Verifier	Functional Importance	DID Management Patterns, "Blockchain and Social Media Account Pair", [58]	I need our service to accept and check credentials that come in different formats or protocols.
Compatibility	Verifier	Problem Importance	DID Management Patterns, "Blockchain and Social Media Account Pair", [58]	Rejecting a credential just because it was issued under another standard would be manageable for us.
Interoperability	Verifier	Functional Importance	DID Management Patterns, "Blockchain and Social Media Account Pair", [58]	I need to verify credentials that come from many different platforms or blockchains.
Interoperability	Verifier	Problem Importance	DID Management Patterns, "Blockchain and Social Media Account Pair", [58]	Accepting credentials from only one ecosystem is sufficient for our service.
Protection	Verifier	Functional Importance	Key Management Patterns, "Hot and Cold Wallet Storage", [58]	I need to verify identity data over a channel that keeps it safe from misuse or tampering.
Protection	Verifier	Problem Importance	Key Management Patterns, "Hot and Cold Wallet Storage", [58]	I would maintain trust in the system even if identity data could be intercepted or altered during verification.
Standard	Verifier	Functional Importance	Trusted Registries category, "Trusted Schemas Registry", [97]	I prefer to accept only credentials that follow recognized standards and schemas.
Standard	Verifier	Problem Importance	Trusted Registries category, "Trusted Schemas Registry", [97]	I would be fine accepting credentials that fail to meet the required standards and schemas.
Transparency	Verifier	Functional Importance	Trusted Registries category, "Blockchain Anchor", [97] Credential Design Patterns, "Blockchain Anchor", [58]	I need to see a clear, unalterable record of how and when credentials are verified.

NFR	Role	Item Type	Design Pattern	Survey Item
Transparency	Verifier	Problem Importance	Trusted Registries category, "Blockchain Anchor", [97] Credential Design Patterns, "Blockchain Anchor", [58]	An audit trail of credential checks feels nonessential to me.
Accessibility	Verifier	Functional Importance	DID Management Patterns, "Identifier Registry", [58] Trusted Registries category, "DID Registry", [97]	I need to access issuer information instantly whenever I verify a credential.
Accessibility	Verifier	Problem Importance	DID Management Patterns, "Identifier Registry", [58] Trusted Registries category, "DID Registry", [97]	I can verify a credential effectively without retrieving issuer details.
Cost	Verifier	Functional Importance	Credential Design Patterns, "Blockchain Anchor", [58]	I need to verify credentials without incurring high transaction or storage fees.
Cost	Verifier	Problem Importance	Credential Design Patterns, "Blockchain Anchor", [58]	I am comfortable paying high costs to verify credentials.
Decentralization	Verifier	Functional Importance	DID Management Patterns, "Identifier Registry", [58]	I need to verify credentials without relying on any single service.
Decentralization	Verifier	Problem Importance	DID Management Patterns, "Identifier Registry", [58]	I am comfortable relying on a single provider to check credentials.
Security	Verifier	Functional Importance	Credential Design Patterns, "Blockchain Anchor", [58] Trusted Registries category, "Status Registry", [97]	I need confidence that security threats to identity data are effectively managed during verification.
Security	Verifier	Problem Importance	Credential Design Patterns, "Blockchain Anchor", [58] Trusted Registries category, "Status Registry", [97]	Effective security management during verification feels optional to me.

Table A.3: Operationalized NFR items for Issuers

NFR	Role	Item Type	Design Pattern	Survey Item
Authenticity	Issuer	Functional Importance	Verifiable Credentials and Presentations category, "Verifiable Attestation", [97]	The ability to digitally sign each credential I issue to guarantee its authenticity and integrity is important to me.
Authenticity	Issuer	Problem Importance	Credential Design Patterns, "Blockchain Anchor", [58]	Issuing credentials without a verifiable digital signature would be acceptable to me.
Compatibility	Issuer	Functional Importance	Trusted Registries category, "Trusted Schemas Registry", [97]	I need to issue credentials that work across both modern identity platforms and traditional identity systems.
Compatibility	Issuer	Problem Importance	Trusted Registries category, "Trusted Schemas Registry", [97]	Credentials that are incompatible with existing legacy systems would be acceptable.
Consent	Issuer	Functional Importance	Verifiable Credentials and Presentations, "Time-Constrained Access", [97]	I need to obtain explicit permission from the holder before issuing any credential about them.
Consent	Issuer	Problem Importance	Verifiable Credentials and Presentations, "Time-Constrained Access", [97]	Issuing a credential without the holder's clear consent would be tolerable.
Interoperability	Issuer	Functional Importance	Trusted Registries category, "Trusted Schemas Registry", [97]	I want the credentials I issue to work with any system or service, no matter the platform.
Interoperability	Issuer	Problem Importance	Trusted Registries category, "Trusted Schemas Registry", [97]	It would be manageable if some services rejected the credentials I issue.
Standard	Issuer	Functional Importance	Trusted Registries category, "Trusted Schemas Registry", [97]	I need to define and publish my credential schemas in widely accepted formats so that any system can process the credentials I issue.
Standard	Issuer	Problem Importance	Trusted Registries category, "Trusted Schemas Registry", [97]	I could tolerate credentials with nonstandard or outdated schemas being unusable by many services.

NFR	Role	Item Type	Design Pattern	Survey Item
Privacy	Issuer	Functional Importance	Verifiable Credentials and Presentations, ”Selective Content Generation”, [97] Credential Design Patterns, ”Selective Content Generation”, [58]	I need to include only the personal data strictly necessary when issuing a credential.
Privacy	Issuer	Problem Importance	Verifiable Credentials and Presentations, ”Selective Content Generation”, [97] Credential Design Patterns, ”Selective Content Generation”, [58]	Including any unnecessary personal data in a credential would be acceptable.
Protection	Issuer	Functional Importance	Key Management Patterns, ”Hot and Cold Wallet Storage”, ”Key Shards”, [58]	I need identity data to be safeguarded from misuse during the credential-issuance process.
Protection	Issuer	Problem Importance	Key Management Patterns, ”Hot and Cold Wallet Storage”, ”Key Shards”, [58]	The possibility of credentials I issue being intercepted or tampered with would be an acceptable risk.
Transparency	Issuer	Functional Importance	Credential Design Patterns, ”Blockchain Anchor”, [58]	I need to record each credential I issue on a public, tamper-proof ledger so anyone can audit my actions.
Transparency	Issuer	Problem Importance	Credential Design Patterns, ”Blockchain Anchor”, [58]	A lack of a tamper-proof record for the credentials I issue would be an acceptable situation.
Verifiability	Issuer	Functional Importance	Verifiable Credentials and Presentations category, ”Verifiable Attestation”, [97]	I must digitally sign each credential and publish its public key or status so it can be independently verified.
Verifiability	Issuer	Problem Importance	Verifiable Credentials and Presentations category, ”Verifiable Attestation”, [97]	I would remain confident in the credentials I issue even if they lack a verifiable signature or current status.
Cost	Issuer	Functional Importance	Wallet and Off-Chain Storage, ”Remote Storage”, [97]	I need to issue credentials without incurring high transaction or storage fees.
Cost	Issuer	Problem Importance	Wallet and Off-Chain Storage, ”Remote Storage”, [97]	A system that charges high fees for issuing credentials would be a minor factor in my decision to use it.
Decentralization	Issuer	Functional Importance	DID Management Patterns, ”Identifier Registry”, [58] Trusted Registries, ”DID Registry”, [97]	I need to issue and revoke credentials using a decentralized registry without relying on any single central authority.
Decentralization	Issuer	Problem Importance	DID Management Patterns, ”Identifier Registry”, [58] Trusted Registries, ”DID Registry”, [97]	I feel comfortable issuing credentials that rely on a single central provider.
Security	Issuer	Functional Importance	Decentralised Identifiers and Cryptographic Keys, ”Key Shards”, [97]	I must ensure that identity data are secure against threats during credential issuance.
Security	Issuer	Problem Importance	Decentralised Identifiers and Cryptographic Keys, ”Key Shards”, [97]	Security measures during credential issuance feel nonessential to me.

A.3 Supplementary Statistical Tests - Identity Holder (Users)

A.3.1 One-Sample Wilcoxon Test: Deviation from Neutral – Identity Holder

Table A.4: One-sample Wilcoxon tests vs. neutral (=3): FI block (Identity Holders, $n = 86$). All items ranked by effect size r (adjusted p).

NFR item	Median	Mean	W	r	95% CI (Mean)	$p_{\text{adj.}}$
Recoverability (NFR16)	5	4.56	3227.5	0.862	[4.41, 4.70]	< 0.001
Security (NFR18)	5	4.56	2956	0.842	[4.39, 4.72]	< 0.001
Persistence (NFR12)	4	4.24	2637	0.828	[4.06, 4.43]	< 0.001
Accessibility (NFR1)	5	4.41	2925.5	0.824	[4.22, 4.59]	< 0.001
Control (NFR7)	5	4.51	3234.5	0.823	[4.33, 4.69]	< 0.001
Protection (NFR15)	5	4.62	2993	0.819	[4.44, 4.80]	< 0.001
Single Source (NFR19)	4.5	4.24	2389	0.802	[4.05, 4.44]	< 0.001
Compatibility (NFR5)	5	4.34	2736	0.799	[4.14, 4.53]	< 0.001
Consent (NFR6)	5	4.23	2451.5	0.798	[4.02, 4.45]	< 0.001
Cost (NFR8)	4	4.17	2445.5	0.794	[3.97, 4.38]	< 0.001
Usability (NFR22)	5	4.36	3021	0.792	[4.17, 4.55]	< 0.001
Privacy (NFR14)	5	4.29	2862	0.787	[4.09, 4.49]	< 0.001
User Experience (NFR23)	4	4.16	2433.5	0.786	[3.95, 4.38]	< 0.001
Authenticity (NFR2)	5	4.33	2858.5	0.785	[4.12, 4.53]	< 0.001
Portability (NFR13)	4	4.20	2922	0.779	[4.00, 4.39]	< 0.001
Transparency (NFR21)	5	4.47	3072.5	0.779	[4.26, 4.67]	< 0.001
Availability (NFR4)	4	4.12	2362.5	0.737	[3.89, 4.34]	< 0.001
Decentralization (NFR9)	4	3.73	1537.5	0.641	[3.50, 3.96]	< 0.001
Autonomy (NFR3)	4	3.77	1852.5	0.632	[3.54, 3.99]	< 0.001
Standard (NFR20)	4	3.86	0.607	1878	[3.60, 4.12]	< 0.001
Interoperability (NFR11)	4	3.65	1804	0.507	[3.40, 3.90]	< 0.001
Existence (NFR10)	4	3.62	1693.5	0.503	[3.35, 3.88]	< 0.001
Non-significant (adjusted)						
Representation (NFR17)	3	3.03	729	0.016	[2.78, 3.29]	0.902
Verifiability (NFR24)	3	3.02	973	0.003	[2.75, 3.29]	0.902

Notes. One-sided Wilcoxon signed-rank tests (alternative = "*greater*"); adjusted p-values within block. 95% CIs for the mean are shown; dashes indicate intervals not reported for non-significant items.

One-sample Wilcoxon signed-rank tests were used to determine whether participants rated each SQRI item above the neutral value of 3 (the midpoint of the Likert scale) and a neutral stance. These one-sided tests assessed if the median rating was greater than 3, indicating above-neutral importance or concern. Analyses were performed separately for the FI and PI_{rev} item blocks, each consisting of 24 statements and based on responses from 86 Identity Holders. To control for multiple comparisons, the Holm adjustment was

applied within each block. Results for each item included the Wilcoxon statistic, raw and adjusted p-values, effect size estimate r, sample median, mean with a 95% confidence interval, and a binomial sign test for robustness. An FI value above 3 indicated higher-than-neutral importance, while a $\mathrm{PI}_{\mathrm{rev}}$ value above 3 indicated that the scenario was at least moderately concerning if the requirement was missing. In other words, $\mathrm{PI}_{\mathrm{rev}}$ scores reflect how problematic respondents find the absence of a given quality or feature. Non-parametric tests were chosen because Likert data are ordinal, may be non-normal, and can show ceiling effects, making them less suitable for parametric analysis.

In the analysis of FI among Identity Holders, 22 of the 24 items were found to exceed the neutral midpoint after Holm correction at $\alpha = 0.05$, with all adjusted p-values falling below 0.001. The remaining two items, *Verifiability* (NFR24) and *Representation* (NFR17), did not yield significant results, as indicated by their adjusted p-values of 0.902 and trivial effect sizes ($r = 0.003$ and $r = 0.016$, respectively). For the significant items, the adjusted p-values ranged from (7.42×10^{-15}) to (4.22×10^{-5}), with effect sizes spanning from $r = 0.003$ to $r = 0.862$. Notably, the strongest effects were observed for *Recoverability* (NFR16; $r = 0.862$, median = 5, $M = 4.56$), *Security* (NFR18; $r = 0.842$, median = 5, $M = 4.56$), *Persistence* (NFR12; $r = 0.828$, median = 4, $M = 4.24$), *Accessibility* (NFR1; $r = 0.824$, median = 5, $M = 4.41$), and *Control* (NFR7; $r = 0.823$, median = 5, $M = 4.51$). The majority of significant items exhibited large effect sizes ($r \geq 0.50$), confirming a strong consensus on their importance to Users, with medians concentrated at 4 or 5 and only two items positioned at the neutral median of 3. A binomial sign test corroborated these findings, reinforcing the non-significance of the two items.

The results from the $\mathrm{PI}_{\mathrm{rev}}$ block concerning Identity Holders highlight notable findings in the area of perceived Implementation Quality. Out of 24 evaluated items, 19 were rated significantly above the neutral midpoint once Holm correction was applied at an α level of 0.05. The adjusted p-values for these significant items ranged from 1.20×10^{-15} to 9.37×10^{-4}. Effect sizes varied widely, ranging from $r = 0.159$ to $r = 0.838$, with a mean effect size of $\bar{r} = 0.57$ and a median of $r = 0.66$. Most of the significant items fell within the medium to large effect size range. Among the items with the strongest effects were *Protection* (NFR15), with an effect size of $r = 0.838$ and $M = 4.70$, median = 5; *Authenticity* (NFR2) at $r = 0.775$, $M = 4.26$, median = 5; and *Cost* (NFR8) at $r = 0.753$, $M = 4.08$, median = 4. Other strongly rated items included *Standard* (NFR20; $r = 0.727$, $M = 4.16$, median = 5), *Consent* (NFR6; $r = 0.725$, $M = 4.38$, median = 5), *Control* (NFR7; $r = 0.722$, $M = 4.23$, median = 5), and *Single Source* (NFR19; $r = 0.720$, $M = 4.08$, median = 5). Overall, the medians of the items in the $\mathrm{PI}_{\mathrm{rev}}$ block indicated higher values, with six items rated as 5, 13 items as 4, and 5 items as 3.

Conversely, five items did not show significant differences from neutral ratings post-Holm correction: *Decentralization* (NFR9; $M = 2.83$, median = 3, $p_{\mathrm{adj.}} = 1.00$), *Representation* (NFR17; $M = 2.57$, median = 3, $p_{\mathrm{adj.}} = 1.00$), *Verifiability* (NFR24; $M = 3.15$, median = 3, $p_{\mathrm{adj.}} = 0.300$), *Usability* (NFR22; $M = 3.23$, median = 3, $p_{\mathrm{adj.}} = 0.190$), and *Existence* (NFR10; $M = 3.24$, median = 3, $p_{\mathrm{adj.}} = 0.182$). The application of the binomial sign test supported the significance of 20 out of the 24 items. At the same time, non-significant results were observed for *Representation* (NFR17), *Decentralization* (NFR9), *Verifiability* (NFR24), and *Existence* (NFR10), aligning with the Wilcoxon outcomes for 19 items. It is noteworthy that *Decentralization* (NFR9) and *Representation* (NFR17) recorded mean

Table A.5: One-sample Wilcoxon tests vs. neutral (=3): PI_{rev} block (Identity Holders, $n = 86$). All items ranked by effect size r (Holm-adjusted p).

NFR item	**Median**	**Mean**	W	r	**95% CI (Mean)**	$p_{\text{adj.}}$
Protection (NFR15)	5	4.70	0.838	3425	[4.54, 4.86]	< 0.001
Authenticity (NFR2)	5	4.26	0.775	2554.5	[4.03, 4.48]	< 0.001
Cost (NFR8)	4	4.08	0.753	2520.5	[3.87, 4.30]	< 0.001
Standard (NFR20)	5	4.16	0.727	2829	[3.92, 4.41]	< 0.001
Consent (NFR6)	5	4.38	0.725	3046.5	[4.14, 4.63]	< 0.001
Control (NFR7)	5	4.23	0.722	2893.5	[4.01, 4.46]	< 0.001
Single Source (NFR19)	5	4.08	0.720	2272	[3.84, 4.33]	< 0.001
Transparency (NFR21)	4	4.08	0.690	2556	[3.84, 4.32]	< 0.001
Security (NFR18)	4	3.99	0.668	2260	[3.73, 4.24]	< 0.001
Compatibility (NFR5)	4	3.78	0.665	1561.5	[3.55, 4.01]	< 0.001
Privacy (NFR14)	4	3.86	0.662	2127	[3.64, 4.08]	< 0.001
Interoperability (NFR11)	4	3.83	0.659	1771.5	[3.59, 4.07]	< 0.001
Autonomy (NFR3)	4	3.90	0.653	2176	[3.65, 4.14]	< 0.001
Accessibility (NFR1)	4	4.00	0.643	2680	[3.74, 4.26]	< 0.001
Recoverability (NFR16)	4	3.90	0.591	2522.5	[3.63, 4.16]	< 0.001
Portability (NFR13)	4	3.78	0.573	2180	[3.52, 4.04]	< 0.001
User Experience (NFR23)	4	3.79	0.541	2313	[3.52, 4.06]	< 0.001
Persistence (NFR12)	4	3.80	0.511	2204	[3.53, 4.08]	< 0.001
Availability (NFR4)	4	3.47	0.428	1552.5	[3.23, 3.70]	< 0.001
Non-significant (Holm-adjusted)						
Representation (NFR17)	3	2.57	0.423	439	[2.33, 2.81]	1.000
Existence (NFR10)	3	3.24	0.220	1224	[2.97, 3.52]	0.182
Usability (NFR22)	3	3.23	0.217	962.5	[2.98, 3.49]	0.190
Decentralization (NFR9)	3	2.83	0.193	621	[2.57, 3.08]	1.000
Verifiability (NFR24)	3	3.15	0.159	1047	[2.91, 3.40]	0.300

Notes. One-sided Wilcoxon signed-rank tests (alternative = "*greater*"); p values Holm-adjusted within block. 95% CIs for the mean are shown; dashes indicate intervals not reported for non-significant items.

ratings below 3, indicating a tendency toward neutral to negative perceptions, despite one-sided tests being geared towards values above 3.

A.4 Supplementary Statistical Tests - Verifier

A.4.1 One–sample Wilcoxon Test: Deviation from Neutral – Verifier

One-sample Wilcoxon signed-rank tests (one-sided, alternative = "*greater*") were applied to each SQRI item to assess whether median ratings exceeded the neutral anchor of 3. Tests were conducted separately for the FI and PI_{rev} blocks (13 items per block; Verifiers n = 27), with multiplicity controlled within each block using Holm adjustment.

Table A.6: One-sample Wilcoxon tests vs. neutral (=3): FI block (Verifiers, $n = 27$). All 13 NFR items ranked by effect size r.

NFR item	**Median**	**Mean**	r	W	**95% CI (Mean)**	$\boldsymbol{p}_{\text{adj.}}$
Significant items (ranked by effect size r)						
Standard (NFR20)	5	4.26	0.827	246	[3.85, 4.66]	< 0.001
Privacy (NFR14)	5	4.41	0.818	314.5	[3.99, 4.82]	< 0.001
Protection (NFR15)	5	4.26	0.809	265.5	[3.83, 4.69]	< 0.001
Cost (NFR8)	5	4.15	0.793	241	[3.71, 4.58]	< 0.001
Authenticity (NFR2)	4	4.26	0.773	282.5	[3.89, 4.63]	< 0.001
Accessibility (NFR1)	4	3.81	0.765	160	[3.43, 4.20]	< 0.01
Security (NFR18)	4	4.04	0.728	211.5	[3.61, 4.47]	< 0.01
Transparency (NFR21)	4	3.74	0.623	162.5	[3.27, 4.21]	< 0.05
Verifiability (NFR24)	4	3.89	0.605	275	[3.42, 4.36]	< 0.01
Consent (NFR6)	4	3.85	0.595	194	[3.34, 4.36]	< 0.05
Non-significant items (Holm-adjusted)						
Compatibility (NFR5)	4	3.52	0.391	167	[3.02, 4.01]	0.097
Interoperability (NFR11)	4	3.33	0.208	117.5	[2.81, 3.86]	0.351
Decentralization (NFR9)	3	3.11	0.097	105.5	[2.64, 3.58]	0.351

Notes. One-sided Wilcoxon signed-rank tests (alternative = "*greater*"); p values Holm-adjusted within block. 95% CIs for the mean are shown.

In the analysis of Functional Importance among Verifiers, 10 of the 13 items were found to exceed the neutral midpoint after Holm correction at $\alpha = 0.05$, with all adjusted p-values falling below 0.05. The remaining three items, *Decentralization* (NFR9), *Interoperability* (NFR11), and *Compatibility* (NFR5), did not yield significant results. For the significant items, effect sizes ranged from $r = 0.595$ to $r = 0.827$. Notably, the strongest effects were observed for *Standard* (NFR20; $r = 0.827$, median = 5, $M = 4.26$), *Privacy* (NFR14; $r = 0.818$, median = 5, $M = 4.41$), *Protection* (NFR15; $r = 0.809$, median = 5, $M = 4.26$), *Cost* (NFR8; $r = 0.793$, median = 5, $M = 4.15$), and *Authenticity* (NFR2; $r = 0.773$, median = 4, $M = 4.26$). The majority of significant items exhibited large effect sizes ($r \geq 0.50$), confirming a strong consensus on their importance to Verifiers, with medians concentrated at 4 or 5 and only three items positioned at or near the neutral median of 3.

The results from the PI_{rev} block concerning Verifiers highlight notable findings in the area of perceived PI. Out of 13 evaluated items, 7 were rated significantly above the neutral midpoint once Holm correction was applied at an α level of 0.05. Effect sizes for significant items ranged from $r = 0.525$ to $r = 0.853$. Among the items with the strongest effects were *Protection* (NFR15), with an effect size of $r = 0.853$ and $M = 4.30$, median = 5; *Cost* (NFR8) at $r = 0.775$, $M = 3.81$, median = 4; and *Standard* (NFR20) at $r = 0.739$, $M = 3.89$, median = 4. Other strongly rated items included *Security* (NFR18; $r = 0.607$, $M = 3.96$, median = 4), *Authenticity* (NFR2; $r = 0.571$, $M = 3.85$, median = 4), *Consent* (NFR6; $r = 0.525$, $M = 4.00$, median = 5), and *Privacy* (NFR14; $r = 0.531$, $M = 3.78$, median = 4).

Conversely, six items did not show significant differences from neutral ratings post-Holm

Table A.7: One-sample Wilcoxon tests vs. neutral (=3): PI_{rev} block (Verifiers, $n = 27$). All 13 NFR items ranked by effect size r.

NFR item	**Median**	**Mean**	r	W	**95% CI (Mean)**	$p_{adj.}$
Significant items (ranked by effect size r)						
Protection (NFR15)	5	4.30	0.853	228	[3.92, 4.67]	< 0.001
Cost (NFR8)	4	3.81	0.775	144	[3.42, 4.21]	< 0.01
Standard (NFR20)	4	3.89	0.739	193.5	[3.47, 4.30]	< 0.01
Security (NFR18)	4	3.96	0.607	254	[3.47, 4.46]	< 0.05
Authenticity (NFR2)	4	3.85	0.571	228	[3.33, 4.37]	< 0.05
Privacy (NFR14)	4	3.78	0.531	241	[3.21, 4.34]	< 0.05
Consent (NFR6)	5	4.00	0.525	240	[3.43, 4.57]	< 0.05
Non-significant items (Holm-adjusted)						
Accessibility (NFR1)	3	3.44	0.579	99.5	[3.11, 3.78]	0.051
Transparency (NFR21)	4	3.63	0.517	151	[3.15, 4.11]	0.051
Compatibility (NFR5)	4	3.52	0.446	191	[3.07, 3.96]	0.056
Decentralization (NFR9)	3	3.22	0.217	118.5	[2.78, 3.67]	0.480
Verifiability (NFR24)	3	3.19	0.175	126	[2.73, 3.64]	0.480
Interoperability (NFR11)	3	3.15	0.120	108	[2.63, 3.67]	0.480

Notes. One-sided Wilcoxon signed-rank tests (alternative = "*greater*"); p values Holm-adjusted within block. 95% CIs for the mean are shown.

correction: *Interoperability* (NFR11; $M = 3.15$, median = 3, $p_{adj.} = 0.480$), *Verifiability* (NFR24; $M = 3.19$, median = 3, $p_{adj.} = 0.480$), *Decentralization* (NFR9; $M = 3.22$, median = 3, $p_{adj.} = 0.480$), *Accessibility* (NFR1; $M = 3.44$, median = 3, $p_{adj.} = 0.051$), *Transparency* (NFR21; $M = 3.63$, median = 4, $p_{adj.} = 0.051$), and *Compatibility* (NFR5; $M = 3.52$, median = 4, $p_{adj.} = 0.056$). These results suggest that while verifiers perceive most quality attributes as important in principle (FI), only a subset are experienced as pressing problem areas in current verification practice (PI_{rev}).

A.5 Supplementary Statistical Tests - Issuer

A.5.1 One–sample Wilcoxon Test: Deviation from Neutral – Issuer

In the analysis of Functional Importance among Issuers, 10 of the 12 items were found to exceed the neutral midpoint after Holm correction at $\alpha = 0.05$. The remaining two items, *Transparency* (NFR21) and *Decentralization* (NFR9), did not yield significant results, as indicated by their adjusted p-values of 0.090 and effect sizes of $r = 0.249$ and $r = 0.328$, respectively. For the significant items, the adjusted p-values ranged from 1.58×10^{-7} to 9.25×10^{-4}, with effect sizes spanning from $r = 0.595$ to $r = 0.873$. Notably, the strongest effects were observed for *Standard* (NFR20; $r = 0.873$, median = 5, $M = 4.41$), *Protection* (NFR15; $r = 0.872$, median = 5, $M = 4.73$), *Security* (NFR18; $r = 0.872$, median = 5, $M = 4.70$), *Privacy* (NFR14; $r = 0.872$, median = 5, $M = 4.57$), and *Authenticity*

Table A.8: One-sample Wilcoxon tests vs. neutral (=3): FI block (Issuers, $n = 37$). All items ranked by effect size r (Holm-adjusted p).

NFR item	**Median**	**Mean**	r	W	**95% CI (Mean)**	$\boldsymbol{p}_{\text{adj.}}$
Standard (NFR20)	5	4.41	0.873	528	[4.16, 4.65]	< 0.001
Authenticity (NFR2)	5	4.51	0.872	561	[4.28, 4.74]	< 0.001
Privacy (NFR14)	5	4.57	0.872	595	[4.35, 4.78]	< 0.001
Protection (NFR15)	5	4.73	0.872	595	[4.53, 4.93]	< 0.001
Security (NFR18)	5	4.70	0.872	630	[4.51, 4.89]	< 0.001
Cost (NFR8)	5	4.30	0.852	459.5	[4.00, 4.59]	< 0.001
Interoperability (NFR11)	4	4.24	0.843	551.5	[3.99, 4.50]	< 0.001
Verifiability (NFR24)	4	4.08	0.727	454.5	[3.71, 4.45]	< 0.001
Consent (NFR6)	4	3.95	0.598	445	[3.53, 4.36]	< 0.001
Compatibility (NFR5)	4	3.73	0.595	417	[3.40, 4.06]	< 0.001
Non-significant (Holm-adjusted)						
Decentralization (NFR9)	3	3.35	0.328	223.5	[2.96, 3.75]	0.090
Transparency (NFR21)	4	3.41	0.249	279.5	[2.92, 3.89]	0.090

Notes. One-sided Wilcoxon signed-rank tests (alternative = "*greater*"); p values Holm-adjusted within block. 95% CIs for the mean are shown for all items.

(NFR2; $r = 0.872$, median = 5, $M = 4.51$). The majority of significant items exhibited large effect sizes ($r \geq 0.50$), confirming a strong consensus on their importance to issuers, with medians concentrated at 4 or 5 and only one item positioned at the neutral median of 3.

Table A.9: One-sample Wilcoxon tests vs. neutral (=3): PI_{rev} block (Issuers, $n = 37$). All items ranked by effect size r (Holm-adjusted p).

NFR item	**Median**	**Mean**	r	W	**95% CI (Mean)**	$\boldsymbol{p}_{\text{adj.}}$
Security (NFR18)	5	4.84	0.872	666	[4.69, 4.99]	< 0.001
Verifiability (NFR24)	5	4.43	0.841	551	[4.13, 4.73]	< 0.001
Authenticity (NFR2)	4	4.30	0.822	578	[4.02, 4.58]	< 0.001
Protection (NFR15)	5	4.73	0.818	645.5	[4.47, 4.99]	< 0.001
Transparency (NFR21)	5	4.16	0.755	462.5	[3.81, 4.52]	< 0.001
Cost (NFR8)	4	3.92	0.689	416	[3.56, 4.27]	< 0.001
Privacy (NFR14)	4	4.05	0.674	468	[3.67, 4.44]	< 0.001
Consent (NFR6)	4	3.81	0.481	409.5	[3.33, 4.29]	0.011
Non-significant (Holm-adjusted)						
Decentralization (NFR9)	3	3.24	0.289	183.5	[2.90, 3.59]	0.290
Interoperability (NFR11)	3	3.22	0.215	235.5	[2.84, 3.59]	0.358
Compatibility (NFR5)	3	2.86	0.174	110.5	[2.51, 3.22]	1.000
Standard (NFR20)	3	3.00	0.024	226	[2.53, 3.47]	1.000

Notes. One-sided Wilcoxon signed-rank tests (alternative = "*greater*"); p values Holm-adjusted within block. 95% CIs for the mean are shown for all items.

The results from the PI_{rev} block concerning Issuers highlight notable findings in the area of PI. Out of 12 evaluated items, eight were rated significantly above the neutral midpoint once Holm correction was applied at an α level of 0.05. The adjusted p-values for these significant items ranged from 5.76×10^{-8} to 1.11×10^{-2}. Effect sizes varied widely, ranging from $r = 0.481$ to $r = 0.872$. Most of the significant items fell within the medium to large effect size range.

Among the items with the strongest effects were *Security* (NFR18), with an effect size of $r = 0.872$ and $M = 4.84$, median = 5; *Verifiability* (NFR24) at $r = 0.841$, $M = 4.43$, median = 5; and *Authenticity* (NFR2) at $r = 0.822$, $M = 4.30$, median = 4. Other strongly rated items included *Protection* (NFR15; $r = 0.818$, $M = 4.73$, median = 5), *Transparency* (NFR21; $r = 0.755$, $M = 4.16$, median = 5), *Privacy* (NFR14; $r = 0.674$, $M = 4.05$, median = 4), and *Cost* (NFR8; $r = 0.689$, $M = 3.92$, median = 4).

Conversely, four items did not show significant differences from neutral ratings post-Holm correction: *Decentralization* (NFR9; $M = 3.24$, median = 3, $p_{adj.} = 0.290$), *Interoperability* (NFR11; $M = 3.22$, median = 3, $p_{adj.} = 0.358$), *Standard* (NFR20; $M = 3.00$, median = 3, $p_{adj.} = 1.000$), and *Compatibility* (NFR5; $M = 2.86$, median = 3, $p_{adj.} = 1.000$). The application of the binomial sign test supported the significance of 8 out of the 12 items, aligning with the Wilcoxon outcomes. It is noteworthy that *Compatibility* (NFR5) recorded a mean rating below 3, indicating a tendency toward neutral to negative perceptions, despite one-sided tests being geared towards values above 3.

Appendix B

Datasets and GitLab Repository

The datasets and the scripts for the statistical analyses were uploaded to the University of Zurich's GitLab server and provided to the supervisor. Access to the folder must be granted before accessing it.

GitLab Repository: Link to GitLab