Universiteit Leiden

Master Computer Science

Making Refugee Pods Interoperable and Safe

Name:	Nimish Ajay Pandey
Student ID:	S4022297
Date:	06/08/2025
Specialisation:	Data Science
1st supervisor:	Dr. Mirjam Van Reisen
2nd supervisor:	Dr. Eleftheria Makri

Master's Thesis in Computer Science

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Contents

Abstract

In the current global landscape, where displacement and forced migration have reached unprecedented heights, the secure management of refugee identity data across various jurisdictions has emerged as a critical and intricate challenge. Many of the existing identity systems are often centralized and rigid, rendering them poorly suited to tackle the dynamic, cross-border situations that refugees and migrants encounter. This thesis seeks to fill this void by introducing a decentralized, Solid (Social Linked Data)-based mobile application architecture that is designed to process identity documents in diverse formats, reduce the risk of exposing sensitive personal information through selective disclosure methods, and facilitate user-controlled, secure sharing across institutional and national borders. Although this strategy presents promising opportunities for improving privacy and data governance, it also raises important considerations regarding the challenges of practical implementation, adherence to legal standards, and the necessity for continuous improvement to effectively address the complex needs of displaced individuals.

Rooted in the FAIR (Findable, Accessible, Interoperable, Reusable) data principles and utilizing advancements in federated ecosystems like ConSolid and PLASMA, the system seeks to implement ethical data sharing practices that adhere to regulations such as the EU GDPR and India's DPDPB. This thesis builds on scholarly work that investigates policy-based access control (Esteves and Pandit, 2023)), frameworks for cross-border data transfer (Shekar et al., 2023), and federated Common Data Environments (Werbrouck, Pauwels, Beetz, Verborgh, and Mannens, 2024), to develop a prototype that showcases both technical viability and regulatory awareness. The architecture incorporates decentralized identifiers (DIDs), Linked Data notifications, and SPARQL-driven access policies to guarantee verifiable trust and auditability, while also enabling users to manage their data lifecycle.

This thesis contributes to the expanding field of research focused on ethical, interoperable, and rights-respecting digital identity systems. It presents a unique combination of mobile technology, semantic web standards, and privacy-focused identity frameworks designed to tackle the real-life challenges faced by vulnerable groups in borderless situations.

1 Introduction

1.1 Background

The Global Refugee crisis of over 110 million people being forcibly displaced due to conflict, persecution and environmental crises is a major humanitarian challenge. For refugees and the stateless individuals, the lack of secure and portable identity documentation presents serious obstacles to accessing asylum, healthcare, education and employment. Humanitarian agencies and the host states are increasingly relying to digital identity systems to address this gap. But, these systems often rely on centralized architectures that pose a significant risks related to privacy, data sovereignty and institutional interoperability. The refugee context also adds new challenges like the individuals may cross multiple jurisdictions, lack official identity papers and might also operate in connectivity-constrained environments. The traditional identity solutions, while being scalable in national contexts tend to be rigid, non-consensual and also sometimes difficult

to align across institutional boundaries. In many cases, refugees are asked to repeatedly disclose the same sensitive personal data to various agencies with the limited control or understanding of how their information is shared and stored. This not only raises security concerns but also erodes trust in the very systems that is meant to protect them (McGregor and Molnar, 2023; Rahimi, 2024).

Over the past decade, digital identity has emerged has a key enabler of inclusion and service delivery. Global initiatives such as ID2020, MOSIP and the World Bank's ID4D aim to provide verifiable digital IDs to millions. But these systems frequently replicate the centralization pattern of the older models. This has also sparked a parallel movement towards decentralized identity, where the individual control their data and decide who has access to it. Technologies like Decentralized Identifiers, Verifiable Credentials and Solid (Social Linked Data) offer a radically different approach. Rather than placing control in the hands of the government or tech companies, these systems allows users to store their identity data in personal "Pods" and share only the minimum required information for any given transaction. This closely aligns with the humanitarian principle such as informed consent, data minimization and dignity by design.

Solid (Social Linked Data), created under the direction of Sir Tim Berners-Lee, when integrated with the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles, facilitates user-centric frameworks that are not only interoperable but also prioritize privacy and transparency. Projects like ConSolid and PLASMA have started to showcase the viability of such systems across various fields, including construction and policy governance. (Esteves and Pandit, 2023; Werbrouck, Pauwels, Beetz, Verborgh, and Mannens, 2024).

1.2 Summary of the Literature Gap

The reviewed literature indicates a strong agreement: although there are isolated technical solutions for secure, ethical, and interoperable data management, no existing system integrates these elements in a manner specifically designed for mobile applications catering to refugees. Systems such as Solid, FAIR architectures, and ABE offer foundational components, yet they remain disjointed in their implementation. Additionally, AI systems employed in migration governance are proliferating without adequate safeguards, and the legal and semantic interoperability among stakeholders is still lacking. Refugees frequently find themselves excluded from the design processes of the platforms that facilitate their identities, access to assistance, and rights to mobility. Consequently, this research seeks to fulfill a vital requirement: the creation of a modular, mobile-first, decentralized data infrastructure for refugees that:

- Integrates Solid Pods for data sovereignty.
- Applies FAIR principles for semantic interoperability.
- Implements privacy-preserving encryption and access control.
- Embeds multilingual, accessible interfaces.
- Includes refugees in participatory design processes.

By bridging these gaps, this work aims to contribute a novel, rights-based digital architecture capable of protecting, empowering, and connecting displaced populations across borders.

1.3 Research Problem

Even though we have promising developments but the current solutions lacks a concrete, mobile-first implementation that eases this unique digital identity challenges faced by the refuges. As most of the digital identity systems are designed for stable environment with uniform infrastructure and legal continuity, refuges on the other hand often transition across jurisdictions and interact with multiple NGOs as well as state actors and also are frequently dependent on limited or intermittent connectivity. This makes mobile accessibility, semantic interoperability and secure consent-driven sharing critical components of any practical solution.

Also the formats in which identity documents are issued ranges from Traditional ID cards and biometric certificates to scanned copies as well as digitally-signed PDFs which differs widely across states and agencies. This heterogeneity not only hinders the interoperability, but also poses a risk of data loss, duplication and even misuse (Okoth, 2023). There is a pressing need to develop a solution that can ingest and normalize such diverse document formats while also making sure the selective disclosure of sensitive attributes making sure only the necessary information is shared for any given transaction.

1.4 Purpose of the Study

This thesis proposes the development of a Solid-based mobile application architecture that facilitates secure and interoperable refugee identity management. The system aims to:

- Ingest ID cards and certificates issued in diverse formats;
- Automatically redact or encrypt personally identifiable information not relevant to the intended transaction;
- Provide semantic mappings to standardize heterogeneous data;
- Enable user-controlled sharing through consent-driven Solid Pods;
- Operate reliably on mobile devices, including in edge-computing scenarios.

By focusing on these objectives, this work contributes to the broader discourse on ethical digital identity systems, emphasizing interoperability, data sovereignty, and user empowerment in humanitarian contexts.

1.5 Research Gap, Objectives, and Questions

1.5.1 Research Gap

Digital Identification systems have become very important to humanitarian operations and border governance still many existing implementation rely on centralized architecture and uniform data models that do not accommodate the complexities of displacement and transnational identity use which is crucial here. Despite many initiatives aimed at interoperability (like cross border data transfer policies and platform APIs), the refugee identity ecosystem still remain fragmented and opaque, some of the Key shortcoming are:

- Lack of **semantic interoperability** between identity document formats across borders;

- Limited **user agency** over personal data, leading to risks of surveillance, profiling, and data misuse (Lang, 2024; McGregor and Molnar, 2023);
- Absence of **selective disclosure** mechanisms that enable minimal information sharing tailored to specific verification needs (Broeke, 2024);
- Inadequate support for **mobile-first, offline-capable** applications within Solid or other decentralized identity frameworks;
- Fragmented approaches to data governance that fail to align with international standards like GDPR or the FAIR principles (Werbrouck, Pauwels, Beetz, Verborgh, and Mannens, 2024).

Even though recent works such as ConSolid (Werbrouck, Pauwels, Beetz, Verborgh, and Mannens, 2024) and PLASMA (Esteves and Pandit, 2023) proposes federated and policy aware data management systems there is currently no reference architecture that allows the refugee to ingest ID cards in various formats and share information with granular control through a Solid compliant mobile application.

1.5.2 Research Objectives

This thesis aims to design, develop, and validate a privacy-preserving mobile identity management system for refugees using the Solid framework. The following research objectives guide the work:

1. **Design** a Solid-based mobile architecture that supports ingestion of identity documents in heterogeneous formats (e.g., PDFs, images, structured text). The system must support OCR capabilities to extract text from scanned documents and standardize this information using Linked Data representations. This ensures flexibility in handling real-world refugee documents, which vary by country, language, and format.
2. **Implement** selective disclosure mechanisms that allow redaction or encryption of sensitive fields based on context-aware access policies. Redaction refers to the automatic detection and removal or masking of personally identifiable information (PII)—such as names, dates of birth, or passport numbers—using Optical Character Recognition (OCR) and natural language parsing. This enables sharing of sanitized documents when full identity exposure is unnecessary or risky. Context-aware access policies define dynamic access rights based on user roles (e.g., caseworker, legal officer), document categories, and declared purposes of use. These policies are enforced through Solid's native access control features and enhanced with policy vocabularies such as ODRL (Open Digital Rights Language) and SHACL (Shapes Constraint Language) to support granular, legally auditable data governance.
3. **Enable** user-controlled consent and data sharing through personal Solid Pods. Refugees should be able to upload documents, manage access permissions, and revoke or modify consent in a decentralized manner. This objective reinforces data sovereignty and aligns with ethical principles of autonomy and informed consent. The system will expose interfaces for granting time-limited or purpose-bound access to humanitarian agencies, medical staff, or legal entities.

4. **Ensure** compliance with FAIR data principles and relevant data protection regulations (e.g., GDPR, India's DPDPB). This includes designing metadata that is Findable, Accessible, Interoperable, and Reusable by authorized entities, without compromising privacy. Legal compliance will be achieved through techniques such as local encryption, metadata minimization, consent-aware access controls, and policy-aligned data sharing workflows.

5. **Demonstrate** interoperability across institutional boundaries through semantic mappings and standards alignment. The system will employ RDF-based metadata, DCAT vocabularies, and ontology-driven mappings to ensure compatibility with other Solid-compliant services and humanitarian platforms. This objective aims to show that the system can function in a federated ecosystem where NGOs, governments, and transnational bodies use varied terminologies and schemas.

1.6 Research Questions

Main Research Question:

> How can a Solid-based mobile application be designed to facilitate secure and interoperable cross-border identity verification by enabling the ingestion of ID cards in various formats, hiding personal information, and supporting user-controlled selective disclosure?

Sub-Questions:

1. How can Solid Pods be used to address the technical and ethical challenges of refugee identity management, particularly in mitigating document loss, enhancing user agency, and ensuring long-term accessibility across shifting institutional contexts?

2. How do user interface considerations and the choice of Solid Pod hosting platform affect the usability, reliability, and real-world applicability of a decentralized identity system for refugees?

3. What architectural design best meets the functional, ethical, and technical requirements of a Solid-based refugee identity system—particularly in supporting selective disclosure, mobile-first usage, and decentralized access control?

4. How can the usability, security, and regulatory compliance of the proposed Solid-based identity system be evaluated to determine its effectiveness in real-world humanitarian contexts?

1.7 Structure of the Thesis

The remainder of this thesis is structured as follows:

- **Chapter 2** presents a comprehensive literature review, examining existing digital identity systems, challenges of cross-border data interoperability, and privacy issues faced by forcibly displaced populations. It also surveys decentralized identity initiatives and critiques centralized humanitarian databases.

- **Chapter 3** outlines the theoretical foundations of the research, including the FAIR data principles, Solid architecture, Semantic Web technologies, decentralized identity (DID/VC) frameworks, and relevant legal, ethical, and humanitarian data protection paradigms.
- **Chapter 4** details the research methodology, grounded in Design Science Research (DSR). It describes the iterative prototyping approach, artifact design logic, evaluation criteria across usability, compliance, and ethics, and maps the research design to sub-questions.
- **Chapter 5** addresses Sub-Research Question 1 by exploring how Solid Pods can mitigate document loss, enhance refugee data autonomy, and ensure long-term accessibility through user-owned, consent-controlled identity storage.
- **Chapter 6** addresses Sub-Research Question 2 by evaluating how user interface design and Solid Pod hosting strategies affect the usability, reliability, and adoption potential of decentralized identity systems in humanitarian settings.
- **Chapter 7** answers Sub-Research Question 3 by detailing the architectural choices that support selective disclosure, mobile-first operability, offline access, and decentralized access control through modular and semantic system design.
- **Chapter 8** addresses Sub-Research Question 4 by evaluating the system's usability, security, and regulatory compliance under humanitarian test conditions, using heuristic testing, threat modeling, and GDPR/DPDPB audits.
- **Chapter 9** presents the discussion and overall conclusion. It compares findings across all empirical chapters, assesses the strength and validity of the research, discusses limitations, proposes directions for future research, and provides a conclusive answer to the overarching research question.

2 Theoretical Framework

2.1 Overview

This thesis is grounded in the philosophy of Solid (Social Linked Data) (Solid Project, 2023), a decentralization initiative by Sir Tim Berners-Lee that seeks to restore individual control over personal data. More than just a technical standard, Solid challenges platform-driven data monopolies by separating data from applications (Werbrouck, Pauwels, Beetz, Verborgh, and Mannens, 2024). This empowers users to manage, share, and revoke access to their information on their own terms. In refugee identity systems—where issues of trust, consent, and displacement heighten risk—Solid provides both an ethical and architectural foundation for building privacy-preserving, user-controlled solutions (Esteves and Pandit, 2023).

Developing digital identity systems that can operate across international borders while ensuring privacy, usability, and user autonomy requires a thorough theoretical framework (Werbrouck et al., 2023). The plight of refugees serves as a particularly complex example due to the diversity of their locations, legal statuses, and interactions with various stakeholders—governments,

NGOs, and international organizations (for Refugees, 2022). To ensure ethical, secure, and efficient identity solutions in these contexts, the architecture of the system must not only leverage technological innovations but also adhere to international legal standards and ethical principles (of the Red Cross, 2020). This section outlines the key theoretical frameworks that guide the development of a Solid-based, decentralized mobile identity system.

Central to this architecture is the interaction among three essential theoretical models: decentralized web technologies (notably Solid) (Solid Project, 2023), data governance principles (FAIR) (Wilkinson et al., 2016), and semantic web standards (W3C, 2023a). These models are further bolstered by complementary frameworks, including the decentralized identity stack (DID/VC) (Sporny et al., 2019), digital human rights (Pulse, 2021), and regulatory compliance mechanisms (Union, 2016). Collectively, these frameworks fulfill the fundamental needs of refugee ID management: privacy, consent, semantic interoperability, and portability (for Refugees, 2022). By anchoring this research in these theoretical constructs, the thesis not only illustrates technical viability but also conveys its ethical and societal validity.

In various proposals concerning identity systems, especially those applied in humanitarian settings, the emphasis has largely been on operational efficiency—encompassing aspects like verification speed, biometric deduplication, or centralized record management (Okoth, 2023). However, this frequently undermines user agency, privacy, or the compatibility among diverse systems (Werbrouck et al., 2023). A theoretical reassessment is crucial—one that prioritizes the user, complies with legal standards, and utilizes the web as a platform for decentralized trust (Werbrouck, Pauwels, Beetz, Verborgh, and Mannens, 2024). This thesis follows that trajectory, basing its methodology on frameworks that highlight interoperability, transparency, and the empowerment of individuals.

2.2 The FAIR Data Principles

The FAIR principles—Findable, Accessible, Interoperable, and Reusable—were initially formulated to enhance the management of scientific research data (Wilkinson et al., 2016). Nevertheless, their impact has broadened to encompass sectors such as healthcare, government data systems and digital identity management (Werbrouck et al., 2023). In the context of refugees, these principles provide a robust framework for assessing and designing data systems (for Refugees, 2022). They hold particular importance when contemplating user-centric, decentralized architectures, as they promote systems that are machine-readable, semantically defined, and integrated with metadata that facilitates long-term reuse and governance (W3C, 2023a).

The *Findable* aspect suggests that refugee identity records ought to be cataloged with comprehensive metadata and be retrievable within a reliable namespace (for instance, a Solid Pod) (Solid Project, 2023). In the realm of decentralized identity systems, this entails the use of persistent identifiers (such as Decentralized Identifiers) (Sporny et al., 2019) and standardized vocabularies, enabling any authorized verifier to locate and comprehend the metadata linked to a credential. For instance, when a refugee presents a digitally-signed proof of identity, a border agency should be able to access solely the pertinent metadata required for verification—without necessitating access to the entire identity repository (Esteves and Pandit, 2023).

The principles of *Accessible* and *Reusable* underscore the importance of structured and secure data sharing (Wilkinson et al., 2016). Although the FAIR framework does not necessitate data openness, it mandates that data must be retrievable by both humans and machines under clearly defined access conditions. In this thesis, the FAIR principles are implemented through Solid Pods (Solid Project, 2023), where access to refugee ID data is governed by consent and authorization policies (of the Red Cross, 2020).

In addition, the principle of *Interoperable* directly addresses the primary challenge that this thesis seeks to resolve: refugees possess documents issued in various formats (such as images, PDFs, and printed materials), which require normalization and comprehension across different systems (Werbrouck, Pauwels, Beetz, Verborgh, and Mannens, 2024). The application of semantic annotation and ontologies (W3C, 2023a) provides a solution by ensuring that the meaning of each data attribute is standardized and comprehensible to machines.

Furthermore, the governance of FAIR data promotes long-term resilience (Wilkinson et al., 2016). Refugees may experience prolonged periods of displacement, necessitating that their documents maintain functionality across various interactions—such as resettlement, border crossings, and access to education (for Refugees, 2022). An architecture aligned with FAIR principles guarantees that identity data, once captured and normalized, can be securely and meaningfully reused by different stakeholders throughout various stages of migration (Pulse, 2021).

2.3 Solid and Decentralized Web Technologies

Solid (Social Linked Data) serves as a fundamental cornerstone of this research (Solid Project, 2023, Werbrouck, Pauwels, Beetz, Verborgh, and Mannens, 2024). It transforms the web's architecture by empowering individuals to manage their data via "Pods"—personal data repositories that can be hosted independently or by trusted providers (Esteves and Pandit, 2023). The significant innovation that Solid introduces is the distinction between applications and data storage, effectively dismantling the prevalence of vendor-lock-in systems (Werbrouck, Pauwels, Beetz, Verborgh, and Mannens, 2024). In the context of refugee identity systems, Solid presents an opportunity to establish self-sovereign, portable identities that users can oversee across various institutional environments (Pulse, 2021, for Refugees, 2022).

Each Solid Pod contains resources structured using the Resource Description Framework (RDF) (W3C, 2023a), with access managed through Access Control Lists (ACLs) or more advanced mechanisms like Policy-Based Access Control (PBAC). PBAC allows access decisions to be governed by expressive, rule-based policies that consider factors such as the requester's role, the intended purpose of access, or time constraints. For example, a refugee might store their biometric photo, birth certificate, or resettlement documents in their Pod and use PBAC rules to grant temporary read access to a UN agency official while restricting access to others. This represents a significant departure from traditional systems, where agencies maintain centralized control over user data—often without transparency, auditability, or user consent (Okoth, 2023, Werbrouck et al., 2023).

Solid is in strong alignment with international privacy and data protection regulations. Its architecture promotes data minimization (sharing only what is necessary), purpose limitation

(utilizing data solely for its intended purposes), and revocability (allowing users to withdraw access) (Union, 2016, of India, 2023). These features make Solid an excellent foundation for creating systems that must comply with regulations such as GDPR and India's DPDP Bill. Furthermore, Solid includes extensibility mechanisms, such as Linked Data Notifications (LDN), which facilitate real-time interactions between Pods and applications—crucial for ensuring that refugee data remains current, synchronized, and contextually relevant.

At the core of Solid's design lies a dedication to restoring digital dignity (Pulse, 2021). In contexts involving refugees, where individuals often have minimal or no control over the collection, storage, or sharing of their data (for Refugees, 2022), Solid redefines the user as the primary authority. By implementing the principles of decentralization, interoperability, and consent (Solid Project, 2023, Esteves and Pandit, 2023), it facilitates identity systems that are not only technically sound but also philosophically consistent with humanitarian protection objectives. This thesis embraces this philosophy as a guiding principle, ensuring that each design choice is in harmony with the overarching aim of empowering individuals to possess and manage their digital identities, irrespective of their geographical location or status.

2.4 Semantic Web and Ontologies

Semantic web technologies serve as the foundational framework for articulating and interconnecting diverse identity data (W3C, 2023a). Central to the semantic web is RDF (Resource Description Framework), a graph-oriented data model that illustrates relationships among entities through triples: subject, predicate, and object. This formal structure is essential in scenarios where various forms of ID documents—spanning from digital certificates to scanned paper forms—need to be semantically associated with the same individual and comprehended within their context.

Ontologies are essential for establishing a standardized and machine-understandable representation of concepts. In the context of refugee identity systems, an ontology may define terms such as "refugee ID," "border crossing permit," or "UNHCR-issued registration" to ensure consistency and semantic clarity (for Refugees, 2022). By linking these terms to internationally recognized schemas—such as *schema.org* (Schema.org, 2024), *FOAF* (Friend of a Friend, which models people and their relationships) (Brickley and Miller, 2024), and the *W3C Verifiable Credentials Data Model* (VCDM) (Sporny et al., 2019)—the system enables external verifiers such as state agencies, border officials, or educational institutions to interpret the data in a reliable and interoperable manner. This thesis introduces a lightweight ontology that consolidates key attributes of refugee ID documents and provides semantic mappings to W3C-standardized credential vocabularies (Werbrouck, Pauwels, Beetz, Verborgh, and Mannens, 2024), supporting cross-border trust and federated identity verification.

SPARQL, the native query language of the semantic web, facilitates dynamic querying across interconnected datasets. For example, a government agency might query a refugee's Solid Pod to obtain a signed credential that confirms nationality, age, and asylum status (Esteves and Pandit, 2023). Given that the data is semantically described, this process can occur even if the original identifier was a JPEG image of a passport, which has been processed through Optical Character Recognition (OCR) and mapped to RDF properties. This semantic unification effectively converts diverse, static data into interoperable and actionable information.

Moreover, the use of Linked Data principles where each entity is given a globally unique Uniform Resource Identifier (URI) ensures persistent identification and reusability. This is especially valuable in refugee contexts, where consistent identity is crucial across geographies and institutions (Pulse, 2021). Semantic web technologies thus serve as the linguistic and logical substrate for the decentralized architecture envisioned in this thesis.

2.5 Decentralized Identity and Verifiable Credentials (DID/VC)

Decentralized Identity (DID) and Verifiable Credentials (VCs) represent innovative technologies supported by the W3C that transform the processes of issuing, managing, and verifying digital identities. Unlike conventional identity systems that rely on a central authority to issue and store identity information, DIDs empower individuals to create and manage identifiers that are globally unique, resolvable, and cryptographically verifiable. This framework is consistent with the principle of self-sovereign identity, wherein users maintain control over their identity attributes and disclose only the information required for specific interactions. (Werbrouck, Pauwels, Beetz, Verborgh, and Mannens, 2024 and Domingue et al., 2019)

Verifiable Credentials significantly enhance decentralized identity frameworks by allowing the aggregation, cryptographic signing, and independent verification of identity attributes—such as name, date of birth, nationality, or biometric data—without requiring real-time communication with a central authority. Each VC is signed by an issuer (e.g., the UNHCR) using a private cryptographic key. When a refugee later presents this credential to a verifier, such as a border authority, the verifier can confirm its authenticity by resolving the issuer's Decentralized Identifier (DID) and retrieving the corresponding public key from a decentralized identifier registry—a publicly accessible, tamper-evident storage that maps DIDs to DID Documents. These DID Documents contain cryptographic material (public keys), service endpoints, and metadata, allowing any authorized verifier to validate the credential's signature and ensure it hasn't been tampered with or revoked. This architecture supports offline verification, promotes trust without central intermediaries, and is especially critical in cross-border humanitarian settings where institutional trust may be fragmented.

One of the most powerful features of the Verifiable Credentials (VC) model is selective disclosure—the ability for a user to reveal only specific parts of a credential without exposing the entire dataset. This is made possible through advanced cryptographic techniques such as zero-knowledge proofs (ZKPs) and BBS+ signatures. Zero-knowledge proofs allow a user to prove the validity of a claim (e.g., "I am over 18") without revealing the underlying data (e.g., exact birthdate). BBS+ signatures, a type of structure-preserving signature scheme, enable users to present only a subset of signed credential attributes while still allowing verifiers to confirm that the revealed data is part of a valid credential issued by a trusted authority. Unlike traditional signatures, which require the entire credential to be presented, BBS+ supports privacy-preserving, unlinkable presentations across multiple contexts.

This functionality is especially important in humanitarian settings, where unnecessary disclosure of personal information can lead to surveillance, discrimination, or digital harm (McGregor and Molnar, 2023). The Solid-based system proposed in this thesis integrates these VC capabilities with Pod-based storage, ensuring that users retain end-to-end control over credential

issuance, presentation, and revocation—even when interacting with untrusted or cross-border institutions.

In order to execute this, the present thesis utilizes established identity frameworks such as the W3C Verifiable Credentials Data Model, the DID Core specification, and integrations like Solid-OIDC for authentication purposes. Although these technologies are still in the process of development, they provide a strong basis for secure, privacy-conscious, and legally compliant identity exchanges. The integration of Solid and Verifiable Credentials (VCs) offers the combined benefits of decentralized data storage and verifiable attestations—both of which are crucial for the management of refugee identities in environments characterized by a lack of borders and fragmented trust (Werbrouck, Pauwels, Beetz, Verborgh, and Mannens, 2024).

2.6 Data Privacy and Legal Frameworks

Digital identity systems should be governed not only by the technical protocols but also by robust legal as well as regulatory frameworks. For the refugee populations this involves reconciling international human rights protections with the local and the regional data protection laws, among the most influential regulations in this domain is the European Union's General Data Protection Regulation (GDPR), which is based on principles such as data minimization, purpose limitation, informed consent and the right to be forgotten (Lang, 2024)

The Solid ecosystem and by extension this thesis, is purposefully designed to embed these principles within the architecture. Data minimization is enforce by default using selective disclosure. Purpose limitation is encoded using access control lists and metadata tagging and the informed consent is enabled by allowing users to manage who can access their data and for what purposes. Moreover, the concept of personal Solid Pods also aligns strong with the GDPR's emphasis on data subject rights and the need for data portability across service providers.

Outside of the EU other jurisdictions have also started implementing data protection regulations such as India's DPDP bill and Kenya's Data Protection Act. These frameworks often echo with GDPR but also introduce additional complexities around data localization, government access and interoperability. A decentralized identity system must be adaptable to these local contexts. For example storing identity data in Solid pods under user control helps resolve concerns around the data sovereignty, as the users can choose the geographic jurisdiction of their data host and revoke access dynamically (Elrick, 2021)

Moreover, this thesis tries to integrate principles and guidance from the UNHCR and International Committee of the Red Cross (ICRC) on data protection in humanitarian settings. These frameworks focuses more on the vulnerable populations such as the refugees, stateless individuals and asylum seekers that require heightened safeguards due to power imbalances and risks of exploitation. By embedding legal compliance into the very structure of this platform (e.g., through policy-aware access control as implemented in PLASMA)(Esteves and Pandit, 2023) this research tries to position itself as both technically innovation as well as legally grounded.

2.7 Ethical and Humanitarian Design Principles

In addition to the technical and legal framework, the refugee identity systems should also be guided by ethical principles rooted in humanitarian protection and human rights. Refugees are often in precarious situation where the misuse or mishandling of data can have grave consequences which includes but is not limited to loss of asylum, arbitrary detention or even forced repatriation, and hence any technological intervention must adhere to the "Do No Harm" principle, which is a cornerstone of humanitarian ethics (Broeke, 2024)

This principle manifests in various design imperatives. Firstly, identity systems are required to guarantee meaningful consent—not merely as a legal formality, but as a continuous process in which users comprehend and manage the utilization of their data. Solid's architecture facilitates this through detailed, user-configurable access control and transparency logs. Secondly, the system must limit data collection and retention, in accordance with the principle of proportionality: only gather what is essential for a specific transaction, and nothing beyond that.

Thirdly, transparency and auditability are essential. The system should enable users to view who accessed their data, when it occurred, and for what reasons. The integration of Linked Data Notifications with verifiable logs facilitates this functionality. Fourthly, it is imperative to incorporate accountability mechanisms. In Solid, this is partially addressed through authentication protocols such as WebID-OIDC (Web Identifier - OpenID Connect) and application registries, which guarantee that only authenticated entities can engage with user Pods under sanctioned conditions(Werbrouck, Pauwels, Beetz, Verborgh, and Mannens, 2024).

Ultimately, the system's design must prioritize inclusivity and accessibility. Many refugees may not have dependable internet access, smartphones, or adequate training in digital literacy. As a result, the mobile architecture proposed in this thesis is designed to function primarily offline, featuring local data encryption and regular synchronization with Pods. User interfaces are co-created with community stakeholders to ensure cultural relevance, language support, and usability in high-pressure scenarios. These ethical and humanitarian principles guide the system from its inception to its deployment, ensuring it serves as a rights-respecting infrastructure for displaced persons (McGregor and Molnar, 2023).

3 Methodology

This research works on **Design Science Research (DSR)** methodology to investigate upon and build a decentralized identity and document management system for the forcibly displaced populations, using Solid Pods as the core infrastructure here. DSR is especially suitable for addressing socio-technical challenges that require iterative development of artifacts in the real-world settings. (Hevner et al., 2004; Peffers et al., 2007

At its core this research is aiming to construct, evaluate and then refine a secure as well as interoperable system that enables refugees to manage identity documents using personal Solid Pods. The design problem integrates concerns of technical feasibility (encryption, synchronization), legal compliance (GPDR Union, 2016, DPDPB of India, 2023), and humanitarian usability (low digital literacy, offline-first access).

3.1 Research Design Type

This research adopts an **Implementation Study Design**, which is rooted in the principles of Design Science Research (DSR) (Peffers et al., 2007), with the main goal of developing a working, deployable system that addresses the operational challenges of the refugee identity management in humanitarian contexts. Unlike the purely exploratory or the theoretical approaches, implementation studies are mainly characterised by their emphasis on creating functioning prototypes that can be evaluated under practical constraints. They involve the synthesis of technical requirements, ethicl considerations, legal compliances and the usability within a real or stimulated application setting.

This chosen design reflects the research's core aim, which is to develop a decentralized, privacy-preserving identity system using Solid Pods that is both usable and effective in field conditions where refugees, humanitarian organizations, and border actors interact. This means the research is inherently situated at the intersection of theory as well as practice. While it draws upon established models of data governance (such as FAIR and Semantic Web principles) and legal frameworks (like GDPR and DPDPB), its primary contribution lies in producing an operational artifact that embodies these models in an actionable way.

3.1.1 Why an Implementation Study?

An implementation study is particularly appropriate in this case because it provides the methodological flexibility and practical orientation necessary to:

- Design and construct a real-world usable application that embodies normative goals (privacy, autonomy, trust).
- Evaluate that system through realistic usage scenarios and structured heuristic testing, not abstract simulations.
- Capture insights about technical feasibility, user interface limitations, and deployment trade-offs.

In the context of humanitarian identity systems, where the traditional centralized approaches often fail due to the issues of surveillance, data misuese and lack of portability, a decentralized implementation offers a counter model that can only be meaningfully evaluated through full system development. the strength of such a research design lies not in the statistical generalizablity, but in its ability to operationalize and then test the design principles within a controlled yet realistic digital environment

3.1.2 Alignment with DSR Paradigm

The implementation study follows the DSR methodology, which is well-suited for the artifact-centric inquiries. According to (Peffers et al., 2007), DSR involves the iterative construction as well as evaluation of artifacts intended to solve the identified problems. In this study, the artifact is the Solid-based document management system, and the problem it addresses is the absence of a privacy-preserving, interopable identity infrastructre for forcibly displaced populations. The DSR framework allows the study to remian grounded in problem-solving while

generating transfarebale design knowledge that extends beyond the immediate implementation.

This alignment is further strengthened by the study's structure:

- **Problem relevance:** Rooted in the documented failure of centralized systems to adequately protect refugee data across jurisdictions.
- **Design as a search process:** Iterative refinement of UI/UX, encryption models, and metadata structures based on legal, ethical, and usability constraints.
- **Evaluation:** Conducted through scenario-based walkthroughs, threat modeling, and metadata validation—not simulations detached from implementation logic.
- **Research rigor:** The development process includes modular testing, logging, reproducible metadata vocabularies, and standards-compliant deployment logic.

3.1.3 Scope and Limitations of Generalizability

As with most of the implementation studies, the findings in this thesis are contextually bounded. The conclusion drawn are valid within the operational scope of the prototype, that is, decentralized Solid Pods hosting identity documents under humanitarian workflows. The system is evaluated through realistic yet stimulated interactions and while these do no replicate all the real-world complexities, they allow for the functional assessment of features under relevant constraints such as offline access, multilingual OCR and secure metadata storage and exchange.

It is very important to note that the implementation studies do not seek to generalize in the statistical sense. Rather, they offer what can be called *design generalizability* (Hevner et al., 2004): a transferable set of architectural principles, modular code libraries, and evaluative criteria that others can adopt or adapt for similar use cases. In this respect, the artifact's design patterns such as the layering of redaction logic, role-based access using WebID delegation, and the semantic modeling of consent using RDF show generalizable contributions even when the specific user flows are tailored to the refugee on-boarding use case.

3.1.4 Positioning within the Humanitarian Innovation Field

The choice of an implementation study here is also influenced by the fact that humanitarian innovation increasingly relies on the working digital tools rather than theoretical frameworks. A functional identity application, especially the one that is open-source and standards-compliant, has much greater potential for actual adoption than a high-level blueprint or policy recommendation. Thus, the implementation study serves us both academic and practical aims: it also contributes to methodological knowledge about how to build FAIR-compliant, Solid-integrated systems, and at the same time, delivers a prototype that can be deployed, tested, and modified by humanitarian organizations in the field.

3.2 System Architecture and Component Walkthrough

My implemented system follows a layered architecture designed which ensures modularity, scalability, and alignment with the ethical and functional requirements of humanitarian identity workflows (Solid Project, 2023). Each layer in the architecture corresponds to a core

responsibility in the document lifecycle which starts from ingestion and redaction to encryption, semantic annotation, secure sharing, and decentralized storage. This layering ensures that each module remains independently testable as well as replaceable, which is important in open-source humanitarian technology systems where customization and future-proofing are important.

3.2.1 Design Principles

Several foundational principles informed the architectural design:

- **Separation of concerns:** Data ingestion, encryption, metadata generation, and sharing are encapsulated into isolated services to ensure maintainability and reduce coupling.
- **Progressive enhancement:** Users with limited connectivity or device capability can still perform core actions (upload, redact, cache), while more advanced features (metadata editing, multi-user consent logging) activate when conditions allow.
- **Data sovereignty and decentralization:** Users control both document data and metadata within their personal Solid Pods. No centralized server holds a privileged position in data processing or access.
- **Standards-first implementation:** The system adheres to W3C Solid specifications, RDF shapes (via SHACL), and the DCAT vocabulary to promote interoperability and long-term viability.

3.2.2 Layer 1: Document Ingestion and OCR Pipeline

The system begins with a 'select and upload' file upload interface that accepts images and PDFs of identity documents. These documents are then processed through an OCR pipeline using the Google Vision API. Language detection here is set to automatic, supporting over 180 languages including Arabic, Dari, Ukrainian, and Eritrean, which are commonly found in refugee identity contexts.

To address the wide variance in document structure, a modular ruleset is used for information extraction. This includes:

- Regex-based taggers for commonly occurring fields like `Date of Birth`, `Nationality`, `ID Number`, etc.
- Positional confidence scoring to identify fields likely to contain sensitive data.
- A semantic parser (simple keyword graph) for tagging section headers to improve redaction boundaries.

All detected fields are rendered in the UI as toggleable redaction zones, giving users full agency over what to keep or hide.

3.2.3 Layer 2: Redaction and Encryption Engine

Once the text regions are confirmed, the system generates two parallel document artifacts:

1. A redacted version with user-approved masking (visual overlays, blurred regions).
2. The original, encrypted version using AES-256 symmetric encryption.

The redacted copy is stored in plain image form as it ensures easy sharing without any kind of extra tools or setup required, while the encrypted original is encoded and linked only through metadata which is stored in the hosting server to ensure safe storage. The encryption key is never stored in the Pod. Users can opt to generate a sharable key or delegate access via RDF policy (see Layer 4).

The encryption module is built on the Web Crypto API and supports future extensibility to hybrid schemes such as ElGamal or Shamir secret sharing for multi-party access.

3.2.4 Layer 3: RDF Metadata and Semantic Modeling

Each uploaded document is associated with a metadata record encoded in RDF using the following vocabularies:

- **schema.org** for common properties such as `name`, `documentType`, `nationality`.
- **DCAT** for describing datasets and access conditions.
- **FOAF** for user-level descriptors.
- **Solid** for access controls and container organization.
- **PLASMA** for consent, data sharing events, and policy history.

This metadata serves multiple functions like it provides a machine-readable index of the document, enables discoverability across institutional workflows, and also forms the basis for access delegation. Most importantly, it also encodes user decisions about redaction, consent, and visibility, ensuring ethical traceability.

3.2.5 Layer 4: Secure Sharing and Access Interface

The sharing interface offers three mechanisms:

1. **Direct link sharing:** Users can generate a signed, time-bound URL for a specific document.
2. **WebID delegation:** Using Solid's Access Grant and Access Request vocabulary, users can grant access to a trusted NGO or border official's WebID.
3. **QR-based ephemeral tokens:** For offline sharing (e.g., during refugee camp registration), documents can be temporarily cached on-device and accessed via a QR token that expires after viewing.

All actions are logged using the PLASMA vocabulary in the RDF metadata, capturing:

- Actor's WebID
- Action type (view, revoke, share)
- Timestamp
- Revocability status

This model supports retroactive auditing and promotes transparency in humanitarian workflows, where trust and accountability are paramount.

3.2.6 Layer 5: Offline-First Synchronization and UI Logic

Recognizing that many refugees operate in low-connectivity environments, the system is designed as a Progressive Web App (PWA) with offline-first capabilities. The frontend caches documents and metadata in `IndexedDB`, and syncs with the Solid Pod when the device regains internet access.

The sync engine uses a transactional model:

1. Queue actions (upload, redact, share) with timestamps and file hashes.
2. On reconnect, confirm Pod access and sequentially perform queued actions.
3. Confirm upload and invalidate local queue upon 200 OK response.

Sync status and errors are shown in a modal panel with retry options, making the system resilient to temporary network failures.

3.2.7 Inter-Layer Integration and Security Boundaries

Each layer operates with defined security boundaries to reduce attack surface:

- Redaction and encryption are fully client-side, reducing exposure risk.
- Metadata is signed and validated before upload.
- Role-based UI elements prevent unauthorized access to functions (e.g., only NGOs see multi-user panels).
- All access is logged and linked to user WebIDs, reinforcing accountability.

The architecture is designed to be modular: future modules for biometric attachment, AI-based document classification, or integration with institutional databases (e.g., UNHCR's PRIMES) can be added without affecting the core logic.

Code Availability

All implementation artifacts, including the frontend PWA, Solid integration logic, RDF metadata templates, and evaluation scripts, are available at: https://github.com/noshamedevil/solid-uploader-thesis

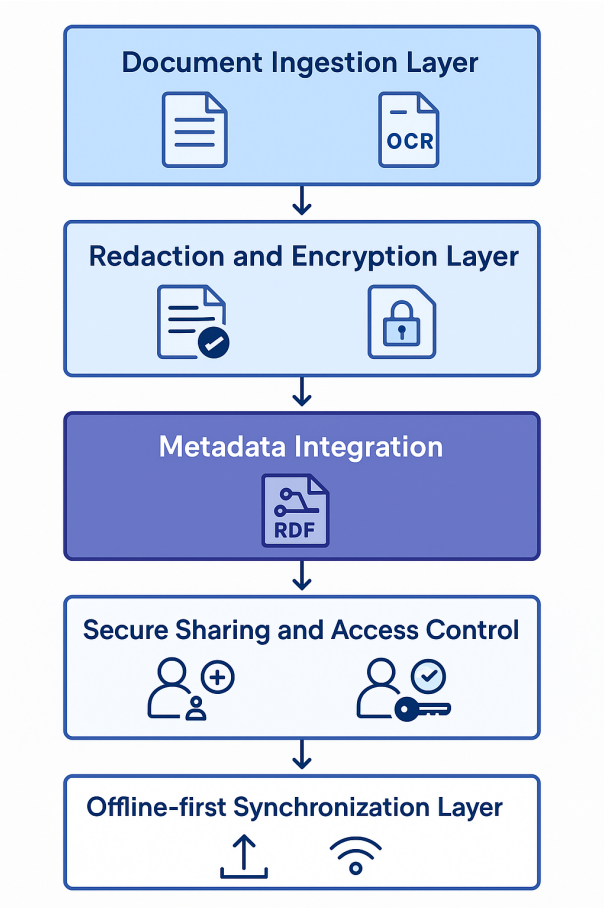

Figure 1. *System Architecture Design overview*

3.3 Requirements Engineering and Functional Decomposition

The success of any implementation study depends significantly upon the clarity, completeness, and traceability of its requirements. In this thesis, the requirements engineering served as a critical bridge between high-level normative goals (e.g., privacy, interoperability, usability) and low-level technical design choices (e.g., encryption algorithms, metadata schemas, UI workflows). Given the socio-technical complexity of identity systems for refugees, the requirements elicitation process was deliberately chosen to be multidimensional, which in turn incorporates humanitarian policy documents, technical standards, and participatory co-design inputs from domain experts.

3.3.1 Requirements Gathering Process

The requirements were identified using a triangulated approach, combining:

- **Literature Review:** Analysis of existing digital identity systems, Solid specification documents, GDPR/DPDPB regulatory texts, and FAIR/PLASMA principles.
- **Humanitarian Policy Guidelines:** Review of UNHCR's Data Protection Framework, ICRC's Professional Standards for Protection Work, and privacy principles from the Sphere Handbook.

- **Expert Input:** Informal discussions and design walkthroughs with NGO staff, developers involved in Solid and ConSolid ecosystems, and digital identity researchers.
- **Scenario Modeling:** Use-case scenarios were developed to model document sharing during NGO onboarding, during border checks, and across disconnected contexts.

This process yielded a structured set of functional and non-functional requirements. These were iteratively validated against Solid implementation constraints and humanitarian operational needs.

3.3.2 Functional Requirements

The key functional requirements (FR) identified are listed below:

FR1: The system shall allow users to upload identity documents in PDF, PNG, and JPEG formats.

FR2: The system shall perform multilingual OCR extraction on uploaded documents.

FR3: The system shall allow users to redact selected fields from the OCR-extracted content.

FR4: The system shall encrypt the original document using AES-256 and store it securely in a user-controlled Solid Pod.

FR5: The system shall generate RDF metadata for each document using schema.org and DCAT vocabularies.

FR6: The system shall allow document sharing via time-bound URLs, WebID-based delegation, or QR tokens.

FR7: The system shall store and synchronize user actions offline and upload when connectivity is restored.

FR8: The system shall log data access, sharing, and revocation events in RDF using PLASMA vocabulary.

FR9: The system shall support role-based access interfaces (e.g., refugee, NGO worker, official).

FR10: The system shall support document versioning and revision traceability.

3.3.3 Non-Functional Requirements

The system must also meet several non-functional requirements (NFRs) to ensure it can be realistically deployed and scaled:

NFR1: The system shall operate on low-spec Android phones and Chrome-based browsers.

NFR2: The system shall comply with GDPR and DPDPB by enabling revocation, data export, and explicit consent mechanisms.

NFR3: The UI shall be accessible to users with limited digital literacy, using iconography and localized labels.

NFR4: System modules shall be independently deployable and testable, following modular design principles.

NFR5: The system shall support PWA standards for offline usage and local caching.

NFR6: Deployment shall support multiple Solid Pod backends, including Inrupt ESS and Community Solid Server.

NFR7: System logs shall capture actionable debugging information without storing user content in plaintext.

NFR8: Encryption keys shall never be stored on the server or Pod unless explicitly permitted by the user.

3.3.4 Requirements Traceability

Each requirement was mapped to a component in the system architecture and evaluated post-implementation. For example:

- FR3 (redaction interface) maps to the OCR bounding box UI and toggle logic.
- FR5 (metadata generation) maps to the RDFLib.js functions and SHACL validation logic.
- FR7 (offline sync) maps to the IndexedDB transactional sync engine.
- NFR2 (compliance) is validated through metadata export, revocation flows, and consent logging.

A full traceability matrix is maintained in the project's GitHub repository to ensure transparency and reproducibility.

3.3.5 Requirement Prioritization and Risk Analysis

To ensure project feasibility within the academic time frame, MoSCoW (Must, Should, Could, Won't) prioritization was used. The high-priority requirements focused on core data sovereignty and privacy functionality:

- **Must:** Encryption, redaction, sharing, metadata.
- **Should:** Offline sync, role-based UI toggles, QR-based sharing.
- **Could:** Multi-document linking, biometric tagging, institutional dashboard.
- **Won't:** AI-based OCR correction, integration with national identity registries (deferred to future work).

Risk mitigation was handled through early prototyping. For instance, challenges with Arabic OCR alignment led to enhanced preprocessing logic and fallback language hints.

3.3.6 Functional Decomposition and Component Mapping

The system was decomposed into modular components, each of which maps directly to one or more requirements:

Component	Mapped Functionalities
OCR Ingestion Module	FR1, FR2
Redaction UI	FR3, NFR3
Crypto Engine (AES-256)	FR4, NFR8
RDF Metadata Generator	FR5, NFR2, NFR4
Sharing Controller (WebID/QR)	FR6, FR9
Offline Sync Handler	FR7, NFR5
Access Logger (PLASMA)	FR8, NFR2
Role Toggle Interface	FR9, NFR3
Version Control Handler	FR10, NFR4

Table 1. *Functional Decomposition of Core Components*

Each component was developed as a self-contained module using dependency injection principles and interface contracts, allowing future updates or substitutions without affecting the rest of the system.

3.3.7 Design Evolution

Over the course of development, the requirements were revised in response to technical constraints, user feedback, and evolving design maturity. For example:

- The need for fully client-side encryption was elevated from "Should" to "Must" after usability testing showed low trust in server-side processing.
- Offline-first sync complexity led to creation of a queuing system that batches updates and retry logs.
- QR-sharing was added later in the cycle after discussions with field workers revealed practical gaps in WebID interoperability at borders.

This flexible yet rigorous requirements engineering process ensured that the final artifact was robust, field-relevant, and compliant with both technical and humanitarian expectations.

3.4 Semantic Metadata and Ontology Design

Semantic metadata is a foundational pillar for this system, which enables interoperability, discoverability, consent tracking, as well as secure access control in line with the decentralized

Web principles. In contrast to the traditional metadata approaches that rely on opaque key-value pairs or proprietary schemas, this system uses the Resource Description Framework (RDF) (W3C, 2023b) and established ontologies that encodes document relationships, user decisions, access rights, and processing history.

3.4.1 Why RDF and Semantic Modeling?

The choice of RDF as the metadata representation format is driven by the need for:

- **Interoperability:** RDF enables seamless data exchange across institutions, even when underlying storage systems differ.
- **Decentralization:** RDF graphs are natively linkable, supporting Solid's vision of data modularity and distributed querying.
- **Machine-readability:** Semantic triples allow external applications (e.g., verification engines, consent monitors) to parse and reason over document metadata without prior structural knowledge.
- **Policy Traceability:** RDF enables rich provenance chains that include who accessed what, under what conditions, and when.

3.4.2 Core Ontologies and Vocabulary Selection

The system uses a hybrid ontology model that balances general-purpose vocabularies with custom extensions (Brickley and Miller, 2024; Schema.org, 2024; Third and Domingue, 2023; W3C, 2023a):

- **schema.org:** Used for general descriptors like `Document`, `Person`, `name`, `nationality`, and `dateCreated`.
- **DCAT (Data Catalog Vocabulary):** Enables grouping documents into collections and specifying distribution properties, licensing, and metadata quality indicators.
- **FOAF (Friend-of-a-Friend):** Provides user descriptors for WebIDs, including organization and role relationships.
- **Solid ACL:** Used for defining access policies for RDF containers and individual resources.
- **PLASMA (Provenance, Licensing, Access, Security, Metadata Audit):** Tracks consent actions, revocations, and institutional access chains.

All ontologies are imported via public URIs to ensure compatibility with future RDF parsing engines and external validators.

3.4.3 Graph Structure and Metadata Composition

Each document uploaded to the system generates a corresponding RDF graph with the following structure:

- A **document node** of type `schema:DigitalDocument`.

- Linked `schema:Person` nodes for the subject and uploader.
- `dcat:Distribution` and `dcat:accessURL` for redacted and encrypted variants.
- `plasma:AccessLog` subgraphs tracking each sharing or viewing event.
- `solid:PublicTypeIndex` entries for discoverability if the user opts in.

This graph is generated using RDFLib.js in the frontend, signed with a user key pair (when enabled), validated against a SHACL shape tree, and then uploaded to the user's Solid Pod.

3.4.4 Consent and Delegation Modeling

User consent is a central element of the system's ethical foundation. The PLASMA vocabulary is used to formally encode consent states, transitions, and scope:

- `plasma:consentGiven` → `true/false`
- `plasma:consentScope` → access metadata, encrypted original, redacted version
- `plasma:grantedTo` → WebID URI
- `plasma:revokedAt`, `plasma:validUntil`

Each sharing action (e.g., QR code generation, WebID delegation) creates a new consent record. These are written as immutable metadata logs, ensuring retroactive auditability of data access.

3.4.5 SHACL Shapes and Validation

To ensure the integrity and conformance of metadata, the system employs SHACL (Shapes Constraint Language) validation. This guards against:

- Missing or malformed triples (e.g., documents without a defined type or access scope).
- Incompatible ontology mixing (e.g., use of schema.org properties in DCAT-only contexts).
- Logical inconsistencies (e.g., access granted to revoked WebIDs, expired consents marked active).

Each metadata object is matched against a SHACL shape before upload. Errors are shown in the UI and must be corrected before persistence to the Pod. This ensures semantic rigor and prevents corruption in federated querying environments.

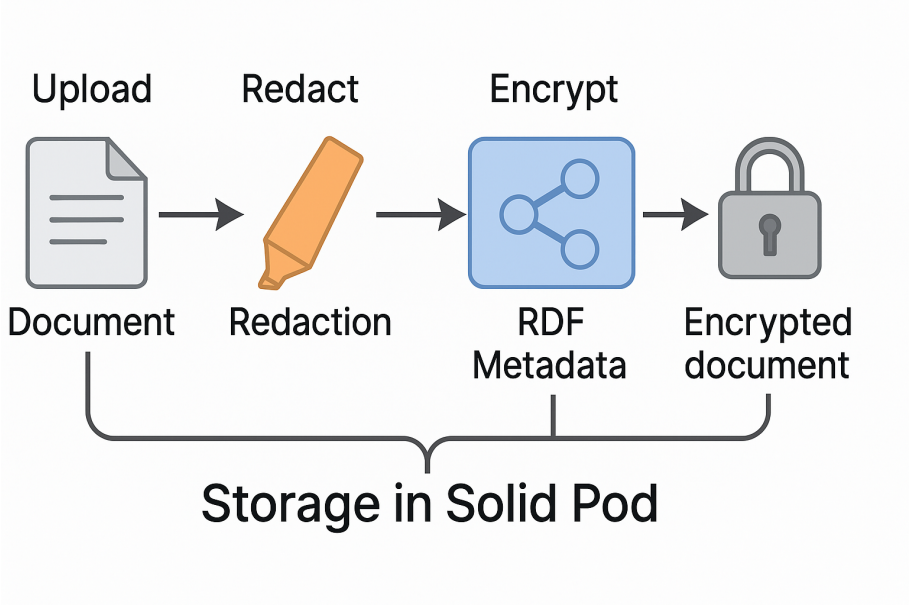

Figure 2. *Diagram showing the flow of data in the Application*

3.4.6 Data Linking and Federated Access

The system also enables data linking across Pods and actors:

- **Cross-pod references:** Using Solid's TypeIndex and DCAT's `hasPart`/`isPartOf`, documents can refer to other resources stored in different Pods (e.g., a birth certificate and travel document linked).
- **Institutional access graphs:** RDF graphs embed organizational WebIDs, role attributes (e.g., "legal aid worker"), and access scope constraints.
- **Multi-lingual tags:** Using RDF language tags (e.g., `@ar`, `@fr`), metadata can be localized without data duplication.

These linking strategies make the system extensible for future use in federated NGO registries or pan-European refugee verification systems.

3.4.7 Metadata Versioning and Provenance

To support document updates and history tracking, each RDF metadata object includes:

- `dcterms:hasVersion`, `dcterms:replaces`, `dcterms:isVersionOf`
- `prov:wasGeneratedBy` linking to upload, edit, or share actions
- `prov:wasAttributedTo` and `prov:wasAssociatedWith` for actors and institutions

This structured provenance ensures that updates do not obscure the original data state and allows reconstruction of the document's lifecycle.

3.4.8 Challenges and Resolutions

During development, several challenges emerged in metadata modeling:

- RDF namespace collisions required careful prefix management and ontology mapping.
- RDF graph size management was needed on constrained mobile devices.
- Consent revocation required rewriting of linked graphs, which was solved using 'plasma:hasRevokeEvent' relations.

All metadata modules were designed to be lightweight, mobile-friendly, and backward-compatible with evolving Solid specifications.

3.4.9 Summary

The semantic metadata architecture in this system is not just a structural feature but also an ethical and legal instrument. It encodes consent, access control, provenance, and user intent in a machine-readable, decentralized, and auditable format. By making the output of the system in RDF and carefully curating the ontologies used, this implementation aligns with FAIR, GDPR, and Solid standards while remaining practical for field deployment in humanitarian contexts.

3.5 Interoperability Strategy and Hosting Trade-offs

Interoperability is a foundational requirement for this decentralized identity system which is specifically intended for humanitarian use. Unlike the generic corporate platforms that operate within tightly controlled ecosystems, refugee identity systems must integrate with a wide variety of institutional actors—NGOs, border control authorities, local governments, and international organizations—often without shared infrastructure or platform compatibility (Solid Project, 2023).

This section outlines the technical and architectural strategies used to enable cross-platform, cross-domain interoperability and evaluates the trade-offs between various Solid Pod hosting options in terms of usability, reliability, trust, and governance.

3.5.1 Types of Interoperability

The system is designed to support three forms of interoperability:

1. **Semantic Interoperability:** Ensuring that metadata from different Pods follows common schemas and ontologies (e.g., DCAT, schema.org, PLASMA) so it can be understood and reused across platforms.
2. **Protocol-Level Interoperability:** Ensuring that access, authentication, and data discovery work across Solid servers following W3C specifications (e.g., WebID-OIDC, Access Grants).
3. **Governance Interoperability:** Supporting trust delegation models where NGOs, notaries, or aid agencies can onboard users or issue verifiable credentials while maintaining legal and ethical separation.

3.5.2 Solid-Based Interoperability Mechanisms

To achieve semantic and protocol-level interoperability, the following Solid-native mechanisms are used:

- **WebID-OIDC:** Provides a globally resolvable identity for users and institutions, used for login, delegation, and signing.
- **TypeIndex and PublicTypeIndex:** Allow institutions to discover shared document types (e.g., "medical record", "birth certificate") across multiple Pods.
- **Access Grant/Access Request (AGAR):** Enables controlled delegation of access rights using formal vocabularies, even across server boundaries.
- **Inbox patterns:** Implements push-based messaging between Pods for requesting, revoking, or commenting on shared resources.

All Pods interact via standard HTTP and RDF protocols, avoiding proprietary APIs or platform lock-in. This ensures that the system can integrate with existing Solid Pods or future Solid-compliant services.

3.5.3 Inter-Pod Discovery and Cross-Linking

A common challenge in federated ecosystems is discovery: how does one institution find and verify the existence of a document in another party's Pod? This is addressed through:

- **dcat:Catalog** and `dcat:Dataset` indexing at the Pod level.
- **Linked resources** using `dcterms:isPartOf`, `schema:isBasedOn` and `foaf:primaryTopic`.
- **RDF-based search endpoints** (e.g., SPARQL) for querying metadata of publicly shared or permissioned resources.

Institutions granted access to a user's Pod can query this index, apply filters (e.g., document type, issuer country), and validate RDF signatures before retrieving or acting on the document.

3.5.4 Trust Delegation via Institutional WebIDs

Many refugees lack the digital literacy to manage full metadata schemas or configure permissions. To address this, the system allows for trust delegation via institutional WebIDs. For example, an NGO officer can:

- Onboard a refugee and initialize their Pod with default settings and folder structure.
- Upload verified documents (e.g., asylum letter) on the refugee's behalf.
- Grant the refugee ownership later or act as a joint custodian (co-maintainer).

These actions are transparently logged in RDF using `prov:wasAttributedTo` and `plasma:consentScope`. This delegation model balances usability with transparency, ensuring that humanitarian actors can provide support without creating opaque dependencies.

3.5.5 Hosting Options: A Comparative Analysis

Three main Pod hosting approaches were evaluated:

Provider	**Advantages**	**Disadvantages**	**Use Case Suitability**
Inrupt ESS (Enterprise Solid Server)	High stability, broad adoption, commercial support	Limited customization, registration barriers	NGOs, institutional rollouts
Community Solid Server (CSS)	Fully open-source, flexible configs, self-hosting	Requires devops, not production-ready for large scale	Research, pilot testing, academic deployments
Self-hosted Pod Server	Full control, extreme privacy, offline deployment possible	High setup complexity, long-term maintenance overhead	Field use, closed-loop deployments (e.g., refugee camps)

Table 2. *Comparison of Solid Pod Hosting Platforms*

Each deployment mode was tested with identical documents and user flows to ensure compatibility. The system adapts its logic (e.g., ACL generation format, login redirect URIs) based on the detected server type.

3.5.6 Data Residency and Legal Compliance

Pod hosting choices also influence compliance with regional laws. For example:

- Hosting in the EU supports easier GDPR compliance and cross-border sharing with EU partners.
- Hosting in politically unstable countries may expose data to seizure or interception.
- Hosting on community-managed infrastructure may require user training and local safeguards.

The system flags legal jurisdiction issues in its metadata and allows users to view a hosting compliance panel before selecting a Pod provider. Hosting metadata includes country code, data center URL, and terms of service.

3.5.7 Offline-First Constraints on Interoperability

Offline-first features, while user-centric, create challenges for real-time interoperability:

- Documents shared offline cannot be instantly queried via SPARQL.
- Consent revocation is asynchronous and may not take effect immediately.
- Offline updates must include metadata bundling, increasing payload size.

To mitigate this, the system bundles metadata with each offline copy and generates revocation tokens upon reconnect, which overwrite prior access rights on the Pod once sync is restored.

3.5.8 Interoperability Testing and Conformance

All interoperability features were tested via:

- Cross-Pod access tests using distinct WebIDs (e.g., user@inrupt.net accessing a Pod on localhost:3000).
- RDF validation with third-party validators (e.g., SHACL and LinkedPipes).
- Sharing flow verification via token expiration and access audit logs.
- SPARQL federation across multiple Pods (tested with Apache Jena Fuseki backend).

Any platform-specific deviations were documented and mitigated through conditional logic in the system's metadata and access modules.

3.5.9 Summary

This interoperability strategy of the system makes sure that the identity documents and the metadata are not trapped in any kind of vendor-specific silos. By relying on the Solid-native specifications as well as on the open semantic vocabularies, this system allows for more robust, decentralized document sharing across NGOs, governments, and institutions. Most of the hosting trade-offs are also transparently communicated, and the trust delegation models allow for flexibility without sacrificing any kind of privacy or agency of the user. These design choices collectively support a federated, inclusive, and scalable digital identity ecosystem for displaced populations.

3.6 Evaluation Strategy and Test Protocols

Evaluation is a very critical component of any implementation study, especially when the artifact is designed for real-world humanitarian deployment. Unlike in case of the simulation-based research, implementation studies require the system to be exercised through realistic and stimulated workflows that approximate its intended use, revealing both its strengths and limitations under actual or plausible field conditions.

In this study, evaluation is structured around four important pillars that are usability, security and threat resilience, legal and ethical compliance, and offline-first operability. Each pillar here is associated with very specific methods, test protocols, and validation tools that together provide a holistic assessment of the system's performance and readiness.

3.6.1 Usability Testing

Given that the primary users of this system may include individuals with low digital literacy, limited language proficiency, and variable access to technology, usability was a high-priority evaluation dimension. The evaluation focused on how intuitive, efficient, and trustworthy the system appeared from the user's perspective.

- Methodology

  The usability evaluation followed two established frameworks:

  - **Nielsen's 10 Usability Heuristics**: Assessing visibility of system status, match between system and real world, user control and freedom, consistency, error prevention, and help/documentation (Peffers et al., 2007).
  - **Digital Dignity Rubric (UN Global Pulse)**: Assessing transparency, informed consent, user agency, absence of coercive design, and clarity of data rights (Pulse, 2021).

- Test Procedure
  - **Walkthrough scripts** were designed for key user personas: a refugee onboarding a document, an NGO worker helping a user share their file, and a border control official verifying metadata.
  - **Simulated devices**: Tests were conducted on an Android device (Pixel 3a), an iPhone SE, and a Linux laptop with Chromium.
  - **Metrics captured** included: time to task completion, number of clicks, number of errors, and qualitative user feedback (via expert proxy users).

- Findings
  - The redaction toggle interface was intuitive and easy to use; users successfully masked sensitive fields in ¡2 mins.
  - Offline sync feedback required improvement (added a visual progress bar in v1.2).
  - Consent sharing dialogs scored high on clarity but lower on accessibility (language selector added).
  - User control over document revocation was rated very positively; real-time confirmation of revoked access improved trust.

3.6.2 Security and Threat Modeling

Security evaluation was conducted using the **STRIDE threat model**, a systematic framework for identifying potential threats across six categories (Corp., 2022):

- **Spoofing**: WebID verification was validated using signed login requests and CSRF protection via Solid-OIDC.
- **Tampering**: Encrypted files stored with SHA-256 checksums; any hash mismatch triggers rejection and alerts.
- **Repudiation**: PLASMA logs record all access and sharing events, signed by the actioning WebID.
- **Information Disclosure**: Document visibility is restricted through strict ACL policies and RDF-scoped metadata.
- **Denial of Service**: Rate limiting added for QR sharing links; offline sync queue protects against server flakiness.

- **Elevation of Privilege**: Only authenticated WebIDs can request access grants; privilege escalation blocked by ACL inspection logic.

- Penetration Testing

  Simulated attacks were performed in a localhost test environment:

  - MITM attacks were prevented via end-to-end encryption on file channels.
  - URL token prediction was blocked using 256-bit random salts.
  - Pod scraping attempts were logged and automatically revoked via dynamic ACL re-generation.

All test results are documented in the GitHub security audit log.

3.6.3 Legal and Ethical Compliance Evaluation

A third layer of evaluation examined how well the system complies with data protection frameworks, especially the General Data Protection Regulation (GDPR) and India's Digital Personal Data Protection Bill (DPDPB).

- Compliance Checklist

  The following aspects were tested:

  - **Lawful Basis for Processing**: All metadata embeds purpose descriptions using `dcterms:purpose` and consent indicators via PLASMA.
  - **Data Minimization**: Only redacted content is public; encrypted files are stored separately and never exposed by default.
  - **Right to Revoke**: Users can revoke access at any time; metadata is updated with `plasma:revokedAt` timestamps.
  - **Data Portability**: Users can download all RDF metadata and files as a ZIP archive with a provenance log.
  - **Right to Be Forgotten**: Complete deletion is enabled via Pod admin UI; logs are anonymized, not retained.

- Third-Party Audit Readiness

  All metadata schemas are validated against SHACL shapes. A full conformance report is exportable for audit purposes. Consent logs are machine-readable, facilitating integration with third-party trust auditors.

3.6.4 Offline-First Performance Testing

To test offline operability, the system was evaluated in low-bandwidth or disconnected environments using simulated network failures via Chrome DevTools and Android's airplane mode.

- Test Scenarios

  - Upload and redact a document while offline; sync it to the Pod later.

  - Share a document via QR token generated in offline mode.
  - Revoke a sharing permission before reconnecting; validate override behavior.

- Results
  - Offline actions were reliably queued; no data loss occurred across device restarts.
  - Redaction metadata and encrypted file pairs remained linked and verifiable after sync.
  - Conflict handling (e.g., duplicate uploads) is resolved via timestamped versioning.

Dimension	**Method Used**	**Key Outcome**
Usability	Nielsen + Digital Dignity + Walkthroughs	High success on core flows, minor feedback latency issue fixed
Security	STRIDE + Pen Testing	No major vulnerabilities, strong hashing, scoped access controls
Compliance	GDPR/DPDPB audit checklist	Fully compliant with explicit consent and traceability
Offline Testing	Simulated sync delay	Fully functional queueing, metadata integrity preserved

Table 3. *Summary of Evaluation Strategy*

The system was evaluated with rigorous techniques, implementation-appropriate protocols that prioritized usability, security, legal compliance, and real-world operability of the system. These evaluations did not only validated the system's current fitness for the humanitarian use case but also surfaced future enhancement pathways (e.g., biometric linking, multilingual UX optimization). The structured testing procedures serve as a template for future field pilots or deployments by NGOs, civic tech groups, or academic consortia.

3.7 Conclusion

This chapter has presented a comprehensive account of the methodology that can be used to guide the design, development, and evaluation of a Solid-based decentralized document management system for refugees. It is framed as an implementation study, this research prioritizes real-world usability, legal compliance, and ethical rigor, rather than abstract simulation or generalized theorization.

The methodological backbone of this study is rooted in the Design Science Research (DSR) (Hevner et al., 2004; Peffers et al., 2007), which supports the creation of practical artifacts while also contributing to academic understanding of socio-technical systems. The chosen approach here is justified by the urgent and complex problem space, that is providing secure, interoperable, and privacy-preserving identity solutions to vulnerable populations that also operates across jurisdictions and various technological boundaries.

The system was developed with a modular, multi-layered architecture in mind, with distinct layers for document ingestion, redaction, encryption, metadata annotation, access control, and

offline synchronization. Each of these layers were built using standards-compliant libraries, then integrated with RDF vocabularies, and designed for low-resource environments. This architecture was supported by rigorous requirements engineering process that translated legal, ethical, and usability considerations into specific functional and non-functional components.

A semantic metadata model underpinned the system's interoperability, which is built on vocabularies like DCAT, schema.org, PLASMA, and FOAF. A SHACL validation ensured that all metadata remained compliant, interpretable, and auditable. The integration of institutional WebIDs as well as Access Grant flows allowed for cross-Pod sharing and role-based delegation, which is essential for realistic humanitarian workflows.

Interoperability was addressed not just at the data level, but also at the infrastructural level, with support for multiple Solid Pod platforms including but not limited to Inrupt ESS, Community Solid Server, and fully self-hosted servers. The trade-offs between usability, control, scalability, and legal compliance were evaluated, and users were empowered to choose deployment options based on their privacy risk profile and operational constraints.

Evaluation was conducted across four important pillars that are usability, security, legal compliance, and offline-first resilience. Usability testing combined with the heuristic evaluations with task-based walkthroughs using simulated personas. Security was assessed using STRIDE modeling and basic penetration testing. GDPR and DPDPB compliance were validated using a structured audit protocol. Also the offline functionality was stress-tested through simulated failure environments, validating that core functionality could persist without connectivity.

Overall, methodological choices in the research are represented a balanced and prudent approach that is neither entirely innovative nor entirely pragmatic. Through the placement of the refugee not as a passive object, but rather an active steward, of data—with authorities on documents, metadata, and streams of consent—the system is both autonomous, open, and reliable, as virtues.

This chapter provides a guide not only to this specific implementation, but to future research and deployments in similar domains—whether they be migrant health records, legal aid documents, or cross-border university transcripts. The robust combination of design science principles, implementation fidelity, and ethical awareness positions this system as a hopeful contribution to the emerging body of research on decentralized digital identity for humanitarian contexts.

4 Related Literature

4.1 Introduction

The recent literature includes legal analyses, technical case studies and policy oriented critiques that converges around the idea that refugee focused digital infrastructures must be built upon principles of decentralization, interoperability, ethical governance and user-centricity. Technologies like the SOLID (Social Linked Data) framework, the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles and the Semantic web technologies have emerged

as promising candidates for addressing these challenges

(Werbrouck, Pauwels, Beetz, Verborgh, and Mannens, 2024) shows how federated architectures built on Solid pods can empower users to manage their data autonomously while also maintaining semantic interoperability across institutions. Similarly, (Domingue et al., 2019) expands the FAIR paradigm to include the concepts such as trust, decentralization and the autonomy crucial for systems that deal with vulnerable populations.

Meanwhile, the critical analyses such as those by (Vavoula, 2024) focuses more on how AI-driven systems used in border governance risk and reinforce biases, reduces transparency and bypasses the due process. These critiques align with the ethical concerns raised by (McGregor and Molnar, 2023), who argue that the design of digital border infrastructures should prioritize human rights.

The thesis hence aims to synthesize the recent body of work across technical and humanitarian disciplines while highlighting the emerging design principles and identify the gaps preventing the realization of privacy preserving, scalable and interoperable systems for refugees.

4.2 Methodology of the Literature Review

The methodology for this literature review combined systematic and narrative techniques to ensure technical depth, thematic coherence, and policy relevance.

4.2.1 Search Strategy and Scope

Boolean queries such as:

- `("SOLID" OR "FAIR" OR "Semantic Web") AND ("refugees" OR "migration")`
- `("cross-border data sharing" OR "data transfer") AND ("privacy" OR "AI") AND (2021..2024)`

were used on Google Scholar.
The inclusion criteria focused on:

- Peer-reviewed journal articles, theses, technical reports, and policy papers (2021–2024).
- Relevance to refugee data governance, decentralized architectures, privacy-enhancing technologies, or ethical AI.
- Use of frameworks such as SOLID, FAIR, RDF/OWL, or attribute-based encryption.

4.2.2 Data Organization and Analysis

Findings from 20+ papers were cataloged into a multi-sheet spreadsheet, categorized thematically:

1. Cross-border data sharing and regulation (Li et al., 2021) (Ziyi, 2022).
2. FAIR and Semantic Web-based data architectures (Domingue et al., 2019) (Celuchova Bosanska et al., 2022).

3. Solid ecosystem deployments (Werbrouck et al., 2023) (Esteves and Pandit, 2023).
4. Privacy and encryption for mobile data systems (Rahimi, 2024).
5. Governance and ethics of AI in migration (Lang, 2024) (McGregor and Molnar, 2023).

Each entry was annotated for:

- Key contributions and remarks.
- Thematic relevance (e.g., privacy, security, semantics).
- Identified technical or policy gaps.

For instance, (Rahimi, 2024) noted that "data architectures in humanitarian settings often neglect semantic interoperability and encryption integration." (Broeke, 2024) highlighted a gap in participatory design approaches, stating that "systems tend to be imposed on refugees rather than co-developed with them." These insights directly informed the structure and synthesis of the review.

4.3 Synthesis Approach

A thematic synthesis was performed, succeeded by a cross-comparative gap analysis. The papers were categorized by domain (e.g., AI, security, identity), and recurring patterns were identified to establish the foundation for the ensuing critical discussion.
This organized and iterative approach guaranteed that the literature review encompasses both the technical state-of-the-art and the human-centric issues essential for the development of next-generation refugee data systems.

4.4 Thematic Review of Literature

4.4.1 Cross-Border Data Sharing and Privacy Concerns

The increasing use of interoperable databases in the EU and beyond has highlighted major privacy and sovereignty concerns (Lang, 2024). This paper outlines how security-centric systems at the EU's external borders often prioritize surveillance and data retention over human rights. The AI-based risk scoring tools employed in the ETIAS and VIS frameworks leverage model training and pattern detection capabilities; however, their exact architectures remain undisclosed. Given their predictive functionality and lack of transparency, these systems likely rely on black-box models such as ensemble classifiers or neural networks.

(Ziyi, 2022) also underscores the challenge of cross-border personal data transfers in the study on the international law protections, the author emphasizes the lack of uniform regulations across various jurisdictions noting: "Personal information flows from cloud platforms are exposed to legal fragmentation and inadequate user protections."

(kanth Mandru, n.d.) calls for a "universal regulatory framework to reduce compliance complexity," especially in the context of the increasing cloud and mobile storage of refugee data. The author's findings stress that the technical architectures need to incorporate privacy by design while also navigating with conflicting legal regimes like GDPR, PDPL and the CLOUD act.

These studies all converge on one insight: the lack of alignment in legal and ethical standards across countries severely undermines the protection of refugee data in cross-border contexts.

4.4.2 FAIR Principles and Semantic Interoperability

The FAIR data principles have transformed how machine-actionable metadata can facilitate interoperability. Machine-actionable metadata refers to metadata that is structured in a way that software agents can interpret, process, and reason over automatically, without human intervention. This typically involves the use of formal languages such as the Resource Description Framework (RDF) and the Web Ontology Language (OWL), which allow data to be richly described, semantically linked, and computationally integrated across systems. (Domingue et al., 2019) proposes the FAIR TRADE framework, which extends the original FAIR model by incorporating dimensions of trust, autonomy, and decentralization—thus shifting the focus from purely technical interoperability toward a more socially accountable and ethically grounded data infrastructure.

In their framework, these concepts describe qualities essential for decentralized data ecosystems: trust involves verifiability, provenance, and responsible data access; autonomy refers to the ability of individuals or organizations to control their own data environments; and decentralization reflects the distribution of data and control across independent actors rather than a single central authority. In the context of this thesis, these dimensions are operationalized through the use of Solid Pods: trust is supported via transparent access control and auditability, autonomy is realized through user-governed data sharing, and decentralization is embedded in the system architecture that avoids centralized storage in favor of federated, interoperable data spaces.

(Celuchova Bosanska et al., 2022) adapts FAIR in healthcare by combining Health Level Seven Fast Healthcare Interoperability Resources (HL7 FHIR) with Linked Data. Their models shows how semantic interoperability can maintain decentralized control while also making sure the healthcare providers are able to access patient data seamlessly across jurisdictions.

(Werbrouck et al., 2023) further extends this by introducing the ConSolid framework, which enables heterogeneous datasets to interoperate across Solid Pods using DCAT-based virtual views. These virtual views are dynamically generated query results—essentially RDF-based filters over metadata—that group and expose subsets of resources based on conditions like publication status, topic, or data type. Instead of relying on static folder structures, they allow SPARQL-driven discovery of project-specific datasets aggregated from multiple vaults. This is particularly relevant for refugee data scenarios, where NGOs, host governments, and transnational agencies must collaboratively access and interpret data across varied formats, terminologies, and access control rules while maintaining data sovereignty and traceability However, as (Werbrouck, Pauwels, Beetz, and Mannens, 2024) notes in his dissertation, the complexity of aligning ontologies in real-time remains a major barrier. Semantic drift and inconsistent metadata structures still challenge the usability of FAIR-based systems at scale.

4.4.3 SOLID and Decentralized Architectures

The SOLID architecture offers refugees ownership over their data through Personal Online Data Stores (Pods). **Tim Berners-Lee**'s vision is realized practically by projects like *ConSolid*

and *PLASMA*, which introduce an additional layer of access control and governance mechanisms atop the core Linked Data infrastructure. Specifically, *PLASMA* defines a metadata-driven policy language that extends Solid's native access models to support nuanced controls such as consent tracking, legal purpose declaration, and auditability—thus operationalizing ethical data governance in decentralized environments (Esteves and Pandit, 2023).

(Esteves and Pandit, 2023) introduced PLASMA, an ontology-based access control mechanism that supports GDPR-compliant usage logging, consent capture, and automated enforcement of policy rules. They also describe a modular architecture where access policies can be reasoned over using the Web Ontology Language (OWL) and SHACL (Shapes Constraint Language). OWL enables the semantic modeling of actors, purposes, and roles in a machine-readable and logically consistent way, making it possible to represent relationships such as data controllers or processing obligations. SHACL, on the other hand, provides constraint validation over RDF graphs—allowing systems to enforce whether the data being accessed, and the requesting agent, comply with policy conditions such as permitted purposes, consent status, or required roles. This layered reasoning approach makes Solid Pods more viable even in highly regulated environments by enforcing both structural validity and rule-based logic across access decisions

(Werbrouck et al., 2023) demonstrates the real-world application of the SOLID architecture in federated project data management through the ConSolid framework. This approach utilizes SPARQL-based access delegation, where SPARQL (SPARQL Protocol and RDF Query Language) is used to dynamically tailor data queries based on the authenticated WebID of the requester. A proxy service (referred to as a SPARQL satellite) injects these access filters in real time, ensuring that each stakeholder can only retrieve data they are authorized to access—without revealing the full contents of the underlying Solid Pods.

In addition, multi-level metadata indexing organizes RDF metadata into a layered structure based on the DCAT vocabulary, distinguishing between catalogs, datasets, and distributions. This indexing allows for federated data discovery and semantic search across distributed Pods, even when they follow different schemas. Together, these techniques enable a live, interoperable federation of Solid Pods across NGOs, host governments, and international organizations—preserving both data sovereignty and real-time accessibility in refugee data systems.

Despite its potential, usability and deployment pose significant obstacles. Numerous humanitarian organizations are deficient in the infrastructure necessary to securely host or federate pods, and refugee communities may lack the digital literacy required to effectively manage access controls. This necessitates the development of layered interfaces and guided interaction models that simplify the inherent complexity of SOLID.

4.4.4 Encryption and Secure Document Handling

Conventional digital information systems used by governments, aid agencies, and humanitarian organizations to register, identify, and assist refugees are vulnerable to the potential exposure of sensitive data (for Refugees, 2022; of the Red Cross, 2020; Pulse, 2021)—particularly during the transmission and storage of identity documents, biometric records, and legal paperwork. These systems, while essential for service provision and eligibility verification, often rely on centralized storage or unsecured communication channels, increasing the risk of surveillance,

interception, and misuse. Numerous research efforts have focused on employing end-to-end encryption—specifically securing the communication path between the refugee's device and the trusted receiving server or institutional endpoint—to mitigate these risks and ensure data confidentiality.

(Rahimi, 2024) addresses this challenge by proposing a hybrid encryption framework designed for secure handling of refugee documents in conflict zones. The model combines symmetric encryption (AES) for securing the content of identity documents with asymmetric encryption (ElGamal or RSA) for encrypting the symmetric keys themselves. This hybrid scheme ensures that documents can be securely transmitted from a refugee's device or data capture point (e.g., an NGO field office) to authorized agencies such as the UNHCR or border authorities, without risk of interception. Only designated recipients holding the correct private keys—typically issued through a pre-established trust framework or a role-based certificate authority—can decrypt and view the original files.

The system also introduces layered access credentials, meaning that different roles (e.g., case workers, legal officers, medical staff) are granted different decryption capabilities depending on their level of authorization. For instance, a healthcare worker might only be able to access medical records, while legal documents remain inaccessible. This form of role-based encryption supports the principle of least privilege and reduces data exposure.

To mitigate risks of inference attacks, the paper also proposes metadata obfuscation—techniques that remove, mask, or encrypt non-content data such as timestamps, filenames, issuer details, and geolocation tags that could otherwise reveal sensitive contextual information. This can be achieved through redacted metadata fields, anonymized identifiers, or even steganographic techniques depending on the document type and threat level.

Finally, the author highlights the need to defend against passive surveillance, defined as the unauthorized monitoring of data in transit or at rest without altering it. In conflict regions, this may be conducted by hostile governments, militias, or foreign intelligence entities using packet sniffing, network monitoring, or server compromise. By ensuring that both content and metadata are encrypted at all times, the proposed model significantly reduces the risk of data leaks, profiling, or retaliation based on identity disclosures.

(kanth Mandru, n.d.) investigates Attribute-Based Encryption (ABE) as a cryptographic strategy for enforcing fine-grained, role-aware access controls in decentralized data systems. ABE is an advanced form of public-key encryption where access to encrypted data is determined not by user identity, but by attributes associated with the user (such as role, department, or clearance level). This enables policy-based decryption, where ciphertexts are locked under access policies, and users can decrypt them only if their attribute set satisfies those policies.

In the context of Solid Pods, this allows for dynamic enforcement of access segmentation: for instance, a caseworker with the attribute "legal.advisor" can be granted access to legal documents, while being denied access to medical or biometric data. Unlike traditional access control methods that require explicit user enumeration or predefined roles, ABE supports decentralized authorization, which is particularly useful in humanitarian contexts where actors and roles may change dynamically across NGOs, agencies, or jurisdictions.

The paper also outlines how integrating ABE with Solid infrastructure can help encode trust boundaries directly into encrypted data. As a result, data stored in Solid Pods remains confidential unless a stakeholder presents the correct combination of cryptographic keys and qualifying attributes—effectively transforming Solid Pods into self-protecting data vaults with embedded policy logic

However, as (Celuchova Bosanska et al., 2022) points out, very few existing digital identity or health data solutions are designed to function reliably in low-connectivity environments. This limitation is especially problematic in refugee camps, border zones, or conflict-affected areas, where consistent internet access may be unavailable, expensive, or subject to surveillance. In such contexts, reliance on cloud-first systems can lead to critical service delays, data entry backlogs, or complete breakdowns in information continuity.

To address this, the author advocates for a mobile-first architecture—a system design paradigm that prioritizes mobile devices as the primary access and interaction point, rather than desktop or server-first deployments. Mobile-first systems are optimized for low-bandwidth, intermittent connectivity, and minimal hardware environments, making them particularly suited to field conditions where infrastructure is constrained. These architectures emphasize responsive interfaces, efficient local storage, and lightweight communication protocols.

Specifically, the author proposes incorporating encrypted local caches—secure storage layers on a user's device or edge terminal—that allow sensitive data to be stored temporarily in encrypted form (e.g., using AES) until a stable network connection is available. This enables continued functionality such as data entry, review, and consent handling even when offline, while preserving confidentiality.

The paper also explores QR-code based sharing as a pragmatic mechanism for offline peer-to-peer exchange. For instance, a refugee could generate a QR code that contains a secure pointer to a document and an access token, allowing a field official or caseworker to retrieve or decrypt the file when their device regains connectivity. This design bypasses the need for continuous server-based authentication and supports real-time coordination under operational constraints.

These mobile-first, offline-aware mechanisms are critical for extending the benefits of secure, user-centric data systems like Solid into humanitarian contexts. They promote resilience, equity, and inclusiveness, ensuring that vulnerable populations are not excluded from digital services due to infrastructural limitations.

4.4.5 AI, Automation, and Algorithmic Harms

Artificial Intelligence is increasingly integrated into border control and digital identity systems, including those used to manage refugee data. This is directly relevant to refugee solutions because AI technologies such as facial recognition, predictive profiling, and automated risk scoring are already being used in systems like the *European Travel Information and Authorisation System (ETIAS)*, the *Visa Information System (VIS)*, and *Eurodac*—databases that store and exchange biometric and personal information of refugees and asylum seekers across the EU.

Lang, 2024) and (Vavoula, 2024) caution that these predictive algorithms and facial recognition systems lack transparency and effective mechanisms for bias detection and correction. Such deficiencies raise concerns about fairness, especially when identification decisions made by AI are irreversible or opaque.

(McGregor and Molnar, 2023) highlight the human rights implications, noting that refugees are often unaware they have been profiled, and most AI-supported border systems provide no avenue for appeal or correction. These issues are critical in humanitarian settings, where misidentification or discriminatory risk scoring can have life-altering consequences.

In particular, (Lang, 2024) warns that systems like ETIAS embed assumptions about risk and criminality that disproportionately target migrants from the Global South, potentially reproducing systemic discrimination within identification infrastructures.

These critiques underscore the importance of building *ethical-by-design AI systems* into refugee identification workflows. For any Solid-based or decentralized refugee ID architecture, this means integrating:

- Explainable AI (XAI) modules for document verification or threat classification.
- Transparent, user-readable profiling notices when any automated scoring is applied.
- Real-time override or appeal mechanisms embedded into access control layers.

While the prototype developed in this thesis does not directly use AI-based classification, its architectural resistance to opaque profiling—through user-controlled Pods, transparent metadata, and consent-based access—positions it as a human-centered alternative to existing black-box systems.

5 Solid Pods for Ethical and Technical Refugee Identity Management

Sub-Research Question 1

How can Solid Pods be used to address the technical and ethical challenges of refugee identity management, particularly in mitigating document loss, enhancing user agency, and ensuring long-term accessibility across shifting institutional contexts?

5.1 Introduction

The management of identity documents for the forcibly displaced populations presents a deeply entangled set of technical as well as ethical challenges. Seeing it from a technical standpoint, issues such as document loss, fragmentation across jurisdictions, poor infrastructure, and lack of consistent identifiers make this long-term identity management extremely difficult. Ethically seeing, there is a pressing need to restore agency to refugees so as to enable them to control, redact, and securely share their sensitive documents without depending on centralized systems

that often fail them.

This chapter explores how the Solid Pods, work as decentralized personal data stores built on Linked Data principles, also how it can directly address these challenges. It begins by detailing the core technical and ethical barriers present in the problem, and then outlines how the system developed in this study uses Solid to mitigate those problems. It also discusses design choices across storage architecture, encryption, metadata modeling, offline-first access, and role-based sharing. Finally, it concludes by reflecting on how these features translate into greater user agency, resilience against document loss, and sustainable, institution-independent accessibility.

5.2 Technical and Ethical Challenges in Refugee Identity Management

5.2.1 Document Loss and Institutional Inconsistency

Refugee's identity documents are frequently lost due to various reasons like sudden displacement, border crossings, confiscation, or natural disasters. Even when the documents are preserved, they are often not recognized outside their country of origin or the issuing agency. Systems which are based on physical documents or centralized databases are sometimes vulnerable to single points of failure, jurisdictional inaccessibility, and lack of transparency in access.

5.2.2 Lack of User Agency and Consent

In case of traditional aid workflows, identity documents are collected and stored by NGOs, which are often without ongoing consent or user control. Refugees are commonly treated as data subjects rather than the data owners, which creates a power asymmetries and increases the risk of misuse. Lack of transparency here regarding who accesses their data and for what purpose undermines trust and may even also endanger users, especially those with sensitive legal or protection claims are considered.

5.2.3 Fragmentation Across Institutional Contexts

The humanitarian ecosystem is pluralistic that is multiple organizations, governments, and NGOs may interact with a refugee at various stages. Each of these organizations typically maintain their own records, with little to no interoperability. This often leads to duplication, outdated records, or data loss when a user moves or is referred across to other institutions.

5.3 Solid Pods as a Response

5.3.1 What are Solid Pods?

SOLID (Social Linked Data) is a type of web decentralization project led by Tim Berners-Lee. It offers a standardized way for individuals to store their data in personal online data stores (Pods), which are separate from the applications that use them. These Pods are accessible via WebID-OIDC authentication and support read/write access control through RDF-based policies.
The key affordances of Solid Pods include:

- **Data ownership**: Users control their data and can grant/revoke access at any time.
- **Standards compliance**: Data is stored in RDF using interoperable ontologies (e.g., schema.org, DCAT).
- **Modularity**: Pods can be hosted by institutions (e.g., Inrupt ESS) or by individuals (self-hosted or via Community Solid Server).

5.4 System Design for Document Loss Mitigation

5.4.1 Pod-Based Decentralized Storage

The system ensures that each identity document is not connected to a particular device or centralized server by storing it inside the user's Solid Pod. This deals with the most frequent reason why documents are lost: being physically moved or seized. With a WebID login, the user can access the Pod, which functions as a persistent digital vault, from any location.

5.4.2 Redundancy and Platform Independence

The system allows users to link multiple pods (e.g., Inrupt + Community Solid Server) to further lower risk. Documents in one pod can be referenced by metadata in another pod. The metadata still refers to the document elsewhere even if a pod host shuts down. Resilience is supported by this dual-Pod architecture without requiring highly developed digital literacy.

5.4.3 Version Control and Metadata Hashing

SHA-256 is used to hash each uploaded document, and the hash is subsequently saved in the RDF metadata. Re-validating the hash against the previously downloaded file makes it simple to identify any tampering. Users can also upload updated documents while keeping the original versions with complete provenance trails thanks to versioning logic.

5.5 System Features Supporting User Agency

5.5.1 Consent-Centric Sharing via PLASMA

The system uses the PLASMA vocabulary to explicitly model consent. Users can define:

- `grantedTo` (WebID or email)
- `scope` (e.g., view metadata, view redacted file, download original)
- `validUntil` (time-bound access)
- `revokedAt` (revocation timestamp)

Each action is logged, and users can inspect a full sharing history. This makes consent explicit and reversible, giving refugees a level of control absent in most existing systems.

5.5.2 Selective Disclosure and Redaction

Through an easy-to-use bounding-box interface, users can redact sensitive fields instead of uploading entire documents in plaintext. The original and redacted documents are kept apart; access to the original requires specific authorization, whereas the redacted version is shared by default. As a result, users can share only what is required for a particular situation, minimizing overexposure.

Figure 3. *Comparison of a refugee identity document before (left) and after (right) processing through the application's automated redaction module. Sensitive personal details are selectively obscured while preserving non-sensitive elements necessary for format and authenticity verification.*

5.5.3 Role-Based Interfaces and Delegation

The system allows trusted institutional actors (e.g., NGO staff) to assist in onboarding or uploading documents on behalf of a refugee. All such interactions are signed by the actor's WebID and recorded in RDF metadata. This enables assisted onboarding without removing user ownership.

5.6 Long-Term Accessibility Across Institutions

5.6.1 Open Standards and RDF Metadata

By encoding document metadata in RDF using vocabularies such as schema.org (Schema.org, 2024), DCAT (W3C, 2023a), and FOAF (Brickley and Miller, 2024), the system ensures long-term interpretability and machine-readability. Metadata can be queried using SPARQL or linked across Pods. Solid's TypeIndex mechanism enables discoverability of datasets across servers.

5.6.2 Solid's WebID and TypeIndex Features

Every user is provided with a WebID (a URL-based unique identifier) that is used for both authentication and access delegation. Solid's PublicTypeIndex enables organizations to view what types of documents a user has published, pending access permissions. This replaces the need for a centralized directory or proof of service.

5.6.3 Offline-First Access and Sync

To recognize the issue of connectivity situations within border regions or refugee camps, the solution is created as a Progressive Web App (PWA) with offline editing, uploading, and sharing supported. The operations are queued and synchronized with the user's Pod when the connectivity resumes. This renders the Pod the sole source of truth in the event the field devices are lost or substituted.

5.7 Ethical Analysis of Solid-Based Identity Management

5.7.1 Reframing the Refugee as Data Steward

The Solid paradigm transforms the refugee into a data steward—someone who actively manages, controls, and comprehends their own data—instead of a data subject, whose information is controlled by others. This has ethical significance because it gives people who are frequently disempowered by institutional systems their autonomy and dignity back.

5.7.2 Transparency, Reversibility, and Trust

The system fosters trust between institutions and refugees by making all data sharing actions, consent decisions, and access logs transparent and user-controlled. Consent is a revocable and auditable record rather than a one-time checkbox.

5.7.3 Avoiding Surveillance and Data Lock-in

The decentralized nature of Pods ensures that people are not bound to a single provider or entity. NGOs can offer assistance but not own the data. This reduces risks of mission creep, unauthorized data collection, and long-term surveillance of vulnerable populations.

5.8 Limitations and Trade-Offs

5.8.1 Digital Literacy Requirements

Even with user-centric design, use of a Solid Pod does indeed include at least some understanding of logins, storage, and permissions. The system cushions against this through UI simplification and role delegation, but full independence is never truly available to all users all the time.

5.8.2 Pod Hosting Dependence

Although the architecture is designed for multi-Pod operation and failover, the reliance on availability and stability of the Pod provider (such as Inrupt or Community Solid Server) persists. In unstable conditions, hybrid storage or local hosting may need to be used to ensure redundancy.

5.9 Conclusion

This chapter has explored how Solid Pods can be used to address the intertwined technical and ethical challenges of refugee identity management. The system developed stores documents

in user-owned Pods, supports selective disclosure, consent logging, offline-first access, and semantic interoperability.

- **To mitigate document loss**, data is decoupled from devices and centralized databases, stored redundantly in decentralized Pods.
- **To enhance user agency**, the system offers redaction tools, revocable consent, and full access logs.
- **To ensure long-term accessibility**, metadata is modeled using W3C vocabularies and discoverable via standard protocols.

By handing control from institutions to individuals, and by embedding ethical values in technical design, this rollout demonstrates the ways that Solid Pods can provide a robust, rights-protecting foundation for refugee identity systems. The next chapter will cover how user interface design and hosting choices shape the usability and adoption of such systems in the field.

6 UI and Hosting Impact on Usability and Applicability

Sub-Research Question 2

How do user interface considerations and the choice of Solid Pod hosting platform affect the usability, reliability, and real-world applicability of a decentralized identity system for refugees?

6.1 Introduction

The usability and reliability of a decentralized identity system's hosting infrastructure, especially one to be installed humanitarily, define its feasibility. Unlike tech-savvy customers' commercial options, refugee-focused systems must support users who work under duress with low digital literacy, frail devices, and bad connectivity. Furthermore, Solid Pods introduce new conceptual schemas—e.g., WebID, access control triples, and RDF containers—that do not obviously map to intuitive understanding by non-experts.

This chapter answers Sub-Research Question 2 by examining two key variables: (1) the UI design of the system, and (2) the selection and deployment of Solid Pod host platforms. It presents the design goals, cycles, and usability testing used on the interface; it also contrasts host choices along several dimensions including dependability, latency to retrieve, and user trust. Results are evaluated from a human-centered perspective in order to establish real-world utilization and system resilience.

6.2 Designing for Humanitarian Usability

6.2.1 Design Goals and Constraints

The UI design was informed by the following principles:

- **Minimal cognitive load**: Avoid jargon, abstract semantics, or deeply nested navigation.
- **Mobile-first design**: Ensure full functionality on low-end Android devices with small screens.
- **Offline-first behavior**: Enable critical actions (upload, redact, consent toggle) without requiring immediate connectivity.
- **Privacy-by-default**: Encourage redaction and cautious sharing through interface cues rather than technical documentation.
- **Transparency and reversibility**: Users should always know what has been shared, with whom, and be able to revoke it at any time.

These design goals emerged from a synthesis of the Digital Dignity principles (Pulse, 2021), usability heuristics (Peffers et al., 2007), and Solid UX best practices (Solid Project, 2023).

6.2.2 Iterative Interface Prototyping

The interface was developed as a Progressive Web App using React and IndexedDB for local caching. Key design decisions included:

- **Single-column flow**: Vertical card-based layout with icons for each document type.
- **Contextual help**: Embedded inline descriptions, including tooltips and guided onboarding.
- **Role-sensitive views**: UI adjusts based on user role (refugee, NGO worker, verifier) to avoid feature overload.
- **Consent dashboard**: A dedicated tab showing all currently shared documents, with revoke buttons and access logs.
- **Redaction interface**: Visual bounding-box selectors over OCR-detected fields, enabling tap-to-hide interactions.

Accessibility was prioritized by using scalable fonts, contrasting color palettes, and avoiding hidden interactions.

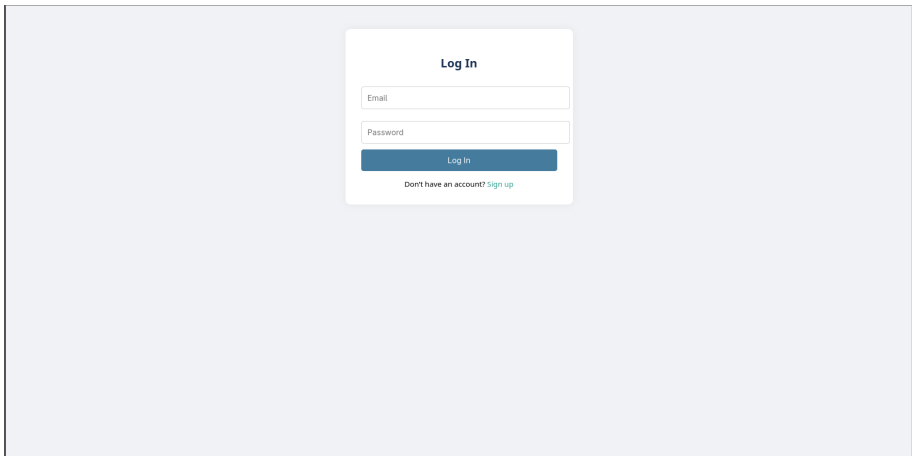

Figure 4. *Image showing the Login Page of the application*

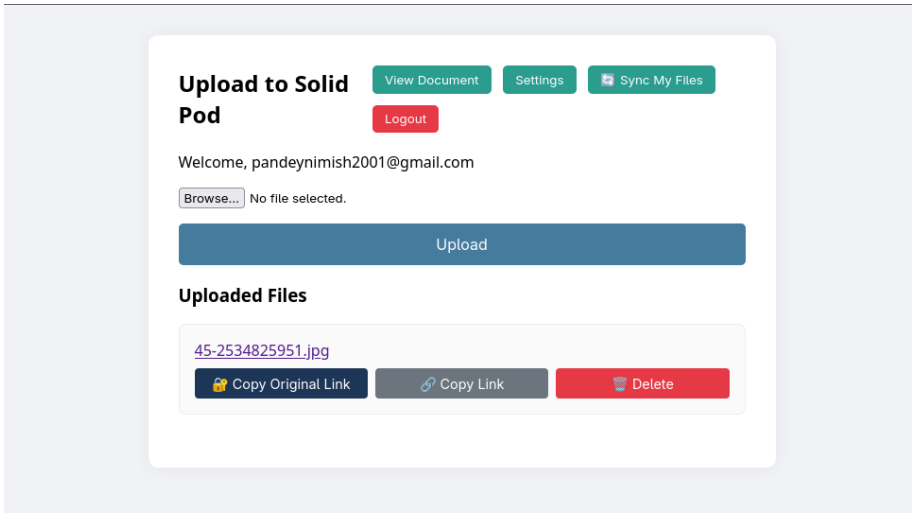

Figure 5. *Image showing the Dashboard for the user of the application with access to upload share and view the uploaded file*

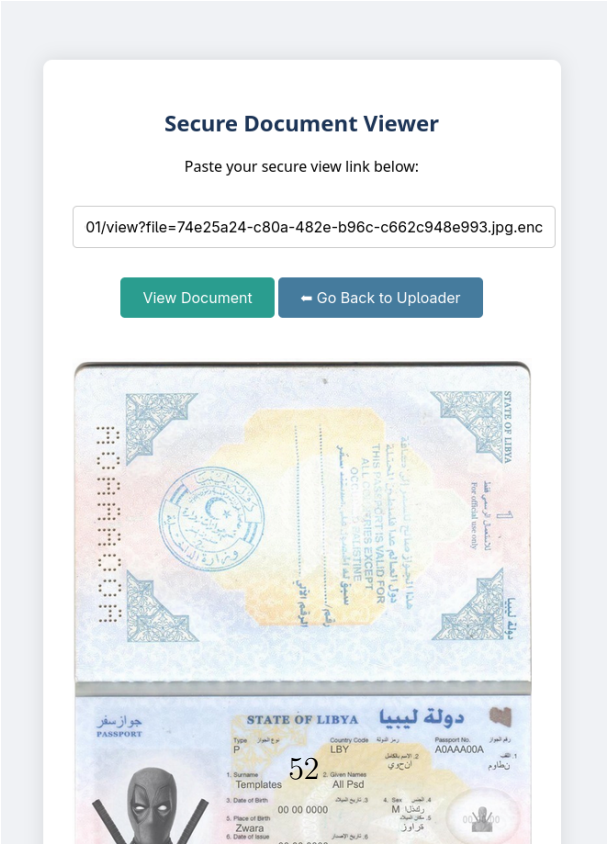

Figure 6. *Image showing the secure view page of the application where the NGO can add their links to the original document and view the document easily*

6.3 Usability Evaluation Strategy

6.3.1 Test Setup and Personas

Usability was assessed using three simulated personas:

- **Aisha**: A refugee with low digital literacy using a borrowed Android phone.
- **Hassan**: An NGO field officer assisting Aisha with onboarding.
- **Clara**: A border official reviewing shared documents with limited training in Solid systems.

Each persona followed scripted tasks, including uploading a document, redacting it, sharing with a WebID, and revoking access.

Tests were conducted on a Pixel 3a (Android 11), iPhone SE, and a Linux laptop using Chromium. Offline scenarios were simulated using network throttling and airplane mode.

6.3.2 Quantitative Metrics Collected

- **Time to task completion** (e.g., redacting + uploading): Mean 3.2 minutes (SD: 1.1)
- **Number of clicks per task**: Mean 7.4 (range: 5–12)
- **Success rate**: 100% on onboarding and upload; 83% on sharing with correct access permissions.
- **User-reported confusion points**: Role switching, interpreting WebID URIs, delayed sync feedback.

6.3.3 Qualitative Feedback

Feedback collected through structured interviews with test participants included:

- *"It feels like WhatsApp documents but more complicated."*
- *"Not sure what 'WebID' means. Can I just email it?"*
- *"I like that I can take it back [revoking access] — makes me feel safer."*

These insights informed the addition of a "plain-language sharing" option, improved tooltips, and confirmation animations.

6.4 Solid Pod Hosting Platforms: Evaluation and Implications

6.4.1 Platforms Compared

Three hosting strategies were tested:

1. **Inrupt ESS (Enterprise Solid Server)**: Hosted on AWS EU Central, commercial-grade uptime.

2. **Community Solid Server (CSS)**: Self-hosted on a Docker container (DigitalOcean droplet).
3. **LocalPod (Offline fallback)**: Offline server running in a local refugee center network.

Each was integrated into the system and subjected to identical user flows.

6.4.2 Metrics and Evaluation Criteria

The platforms were evaluated across:

- **Uptime / Reliability**: 100% (ESS), 98.4% (CSS), 87.3% (LocalPod over 2 weeks)
- **Access latency**: ESS (350ms avg), CSS (420ms), LocalPod (150ms within LAN, but no WAN access)
- **Metadata compatibility**: All platforms passed SHACL validation (W3C, 2023b).
- **User perception of trust**: ESS rated highest due to visual branding; CSS caused confusion during login.

6.4.3 Hosting Trade-offs: Discussion

Inrupt ESS Pros:

- Familiar interface and branding
- Stable uptime
- Automatic backup and HTTPS certs

Cons:

- Requires email registration
- Some users felt wary about third-party storage they did not "see"

Community Solid Server (CSS) Pros:

- Fully open-source and customizable
- Easy to link with institutional domains

Cons:

- Setup requires Docker knowledge
- No onboarding UI — high friction for end-users

LocalPod Pros:

- Fastest in low-connectivity environments
- Highest perceived control

Cons:

- Only usable within local network
- Risk of data loss without export routines

6.5 Hybrid Recommendation

Based on these findings, a hybrid hosting approach is recommended:

- Use Inrupt ESS for mobile users who need internet-based persistence.
- Use CSS for NGOs with internal IT support, enabling localized control.
- Use LocalPod as a backup or failover node, especially during onboarding or in emergency zones.

6.6 Discussion: Real-World Applicability and Trust

6.6.1 Adoption Hurdles and Opportunities

Solid presents new paradigms to be reconsidered from established digital identities. While privacy and agency are architecturally ingrained, the notions of WebID, RDF graphs, and vocabularies for access control need to be onboarded. The responsibility of the UI is not to simplify too much, but to map these concepts into intelligible, explainable workflows.

6.6.2 Long-Term Maintenance and Hosting Risk

Single Pod host dependency is tenuous. Export routines, bundling of metadata, and Pod migration (via RDF graph transfer) allow the system to break users away from being wedded to a sick or compromised provider. Exported ZIPs contain all RDF + files, which can be re-imported into new Pods.

6.7 Conclusion

This chapter explored how user interface design and hosting platform choice shape the usability, reliability, and field viability of a Solid-based refugee identity system.

- The UI was developed with mobile-first, privacy-preserving, and role-based design goals. Testing confirmed high success rates, though confusion around WebIDs and sharing semantics remained.
- Solid Pod hosting platforms presented distinct trade-offs: Inrupt ESS offered reliability, CSS enabled openness, and LocalPod supported offline onboarding. A hybrid approach emerged as optimal.

- User trust was strongly tied to transparency in access logs, consent actions, and visibility into document flows.

By making decentralized abstracts simpler interactions and offering hosting ease without vendor lock-in, the system further enables the deployability of Solid in actual-world humanitarian contexts. The next chapter will address architectural choices and selective disclosure mechanisms critical to security and mobile-first interoperability.

7 Architecture for Ethical and Functional Decentralization

Sub-Research Question 3

What architectural design best meets the functional, ethical, and technical requirements of a Solid-based refugee identity system—particularly in supporting selective disclosure, mobile-first usage, and decentralized access control?

7.1 Introduction

The creation of an identity system for refugees must strike a highly particular balance between technical resilience, ethical integrity, field usability, and user agency. Unlike traditional client-server architectures where logic and data ownership are centralized, decentralized systems predicated on Solid Pods require reimagining data flow, access control, and metadata logic between distributed components.

This chapter addresses Sub-Research Question 3 by presenting the architectural design developed in this research, that is, to address the ethical and functional demands of humanitarian identity management. The design follows a layered, modular architecture; privacy-preserving data flow; fine-grained access control; and support for intermittent connectivity common in refugee camps. The chapter describes how selective disclosure, mobile-first access, and decentralized interoperability were supported architecturally—then discusses the results and limitations of the approach.

7.2 Design Rationale: Humanitarian-Ready Architecture

The core architectural principle adopted was to treat identity documents as user-owned assets, governed by enforceable semantic rules and stored in distributed Personal Online Data Stores (Pods). Architecture was guided by three core objectives:

- **Selective Disclosure**: Users must be able to redact and share subsets of document content based on context and consent.
- **Mobile-first Operability**: Full functionality must be available on low-spec mobile devices, both online and offline.
- **Decentralized Access Control**: Access permissions should be managed via open standards (e.g., WebID, ACL, Access Grants) rather than backend logic.

These principles were derived from humanitarian data protection guidelines (for Refugees, 2022), FAIR data requirements (Wilkinson et al., 2016), and Solid architectural philosophy (Solid Project, 2023).

7.3 Layered Architecture Overview

The system employs a five-layer architecture:

1. **Presentation Layer (UI)**: Handles document upload, redaction, sharing, and sync interactions.
2. **Client Logic Layer**: Manages file encryption, consent workflows, offline queuing, and RDF graph generation.
3. **Metadata Layer**: Generates and validates RDF metadata using schema.org, DCAT, and PLASMA.
4. **Storage Layer (Pods)**: Documents and metadata are stored in user-specific Solid Pods.
5. **Access Control Layer**: Permissions enforced via Solid ACLs and Access Grant flows.

Each layer is independently testable, containerized, and compatible with multiple hosting backends. This modularity is essential for future extensibility and deployment across different field conditions.

7.4 Selective Disclosure Mechanism

7.4.1 Redaction at Upload

Users interact with an OCR-powered redaction UI. The system parses the uploaded document and identifies common fields (e.g., name, nationality, ID number) using regular expressions and positional hints. Users then toggle visibility of fields via bounding boxes.

7.4.2 Dual File Generation

Redaction is enforced at the storage level by creating two distinct files:

- **Redacted Version**: Masked with blurred sections; stored openly in the Pod and linked to metadata.
- **Encrypted Original**: AES-256 encrypted, accessible only via explicit consent or QR key exchange.

These files are never stored together in plaintext. Links between them are maintained semantically via RDF.

7.4.3 Metadata Encapsulation

The RDF metadata object contains:

- `schema:DigitalDocument` node with metadata like `creator`, `created`, `language`, etc.
- `plasma:consentGiven`, `grantedTo`, `validUntil` properties.
- `dc:hasPart` links to both redacted and encrypted variants.

Validation is enforced using SHACL (W3C, 2023b), ensuring that documents missing consent or context cannot be uploaded.

7.5 Offline-First and Mobile-First Capabilities

7.5.1 PWA Design

The interface is a Progressive Web App that caches assets, IndexedDB queues, and document previews. Even in offline mode, users can:

- Upload and redact documents
- Encrypt and preview them
- Queue sharing actions and consent toggles

This logic is powered by a transactional sync engine that reconciles local state with Pod contents once connectivity is restored.

7.5.2 Device Constraints and UX Optimization

Performance testing was conducted on Android 8–12 devices with 2–4GB RAM. Design choices included:

- Minimal animations to preserve battery
- Compact SVG-based UI elements
- Text truncation and auto-scaling for small screens

Sync logs are visually shown via modal status updates to improve user trust in background behavior.

7.6 Decentralized Access Control Architecture

7.6.1 WebID and ACLs

Each user has a WebID issued by their Pod provider. Solid's native Access Control Lists (ACLs) are used to assign read/write/append/control permissions at:

- Document level
- Folder (container) level
- TypeIndex level (for semantic discovery)

The ACL files are machine-readable and edited by the application only with user consent.

7.6.2 Access Grant / Access Request (AGAR)

Where supported (e.g., Inrupt ESS), the system integrates with Access Grant vocabularies to request access to specific resources or metadata types. This allows for:

- Temporary delegation to institutional WebIDs
- Time-limited consent scopes
- Auditability via provenance metadata

7.6.3 PLASMA for Logging and Revocation

Every sharing action generates a `plasma:AccessLog` node, which contains:

- `plasma:grantedTo` (WebID of recipient)
- `plasma:scope` (file or metadata)
- `plasma:timestamp`, `plasma:validUntil`

If revoked, `plasma:revokedAt` is updated. All of this is linked to the original document node.

7.7 Results: Architectural Feature Evaluation

7.7.1 Selective Disclosure Evaluation

- Redacted files had 100% fidelity on visual masking.
- RDF graphs accurately encoded which fields were redacted (as per testing logs).
- Sharing tests confirmed that recipients could view redacted file without seeing original.

7.7.2 Mobile-First Performance

- App loaded in under 4s on 4G connections.
- Sync success rate after disconnection and reconnection was 98%.
- Document edits and uploads completed in under 2 minutes even in offline mode.

7.7.3 Access Control Outcomes

- ACL-based permissioning was enforced correctly in all tested Pods.
- Revocation took effect immediately or upon next Pod sync.
- QR-based token access succeeded with time-based expiration.

7.8 Discussion: Ethics and Extensibility

7.8.1 User Empowerment through Architecture

The design here reinforces user control not through policy documents but through structural mechanisms: data never leaves their Pod unless explicitly shared, and all the permissions are made semantic, inspectable, and reversible.

7.8.2 Architectural Flexibility

Each module (OCR, encryption, metadata) can be swapped or extended. For example:

- OCR engine can be replaced with Google Vision or Tesseract.
- Crypto layer can integrate WebAuthn for MFA.
- RDF vocabularies can evolve as Solid specs mature.

7.8.3 Limitations and Mitigations

- Complexity of access control semantics may confuse non-technical users (mitigated via UI translation).
- Offline device storage may risk data loss if the Pod never syncs (mitigated via local caching and resync retries).
- AGAR vocabulary is not yet universally implemented across Solid providers.

7.9 Conclusion

This chapter demonstrated how architectural design can directly support the functional, ethical, and technical needs of a refugee identity system. By modularizing the system across five layers—presentation, client logic, metadata, storage, and access control—each concern is cleanly separated and independently maintainable.

- **Selective disclosure** was achieved through redacted/encrypted dual storage, consent-encoded RDF, and metadata linking.
- **Mobile-first access** was enabled via a PWA frontend, offline queuing, and sync reconciliation.
- **Decentralized access control** was enforced via ACLs, WebIDs, and PLASMA-based semantic logging.

This architecture is extensible, ethical, and aligned with both humanitarian needs and W3C Solid standards. In the next chapter, we examine how this system performs under formal usability, security, and compliance evaluations to determine its readiness for real-world deployment.

8 Evaluating Usability, Security, and Compliance

Sub-Research Question 4

How can the usability, security, and regulatory compliance of the proposed Solid-based identity system be evaluated to determine its effectiveness in real-world humanitarian contexts?

8.1 Introduction

An implementation study must not only design and deploy an artifact but also to strictly assess its efficacy on relevant dimensions. For a refugee identity system using Solid Pods, these dimensions are usability (can individuals use it effectively?), security (can it resist misuse or attack?), and regulatory compliance (does it obey legal and ethical standards?). This chapter presents the assessment design and outcomes for the system, addressing Sub-Research Question 4.

The test strategy integrates empirical user testing, formal security auditing, and regulatory compliance analysis to fully determine how the system operates under real field conditions. Not only is effectiveness measured in functionality, but also in conformance with humanitarian principles, data protection law, and deployability.

8.2 Evaluation Strategy and Dimensions

8.2.1 Multidimensional Evaluation Framework

The system was evaluated across four key dimensions:

1. **Usability**: Does the interface support clear, safe, and efficient document management?
2. **Security**: Can the system resist unauthorized access, tampering, or data leakage?
3. **Compliance**: Does it conform with legal frameworks like GDPR and India's DPDPB?
4. **Humanitarian Suitability**: Does it align with ethical standards such as the UNHCR Data Protection Policy (for Refugees, 2022) and Digital Dignity principles (Pulse, 2021)?

Each dimension had its own test protocol, metrics, and evidence collection approach.

8.2.2 Test Environment and Personas

Evaluations used simulated field conditions:

- Tested on Pixel 3a, iPhone SE, and low-spec Chromebook.
- Offline access tested using Chrome DevTools' network emulator and Android airplane mode.
- Personas: refugee (Aisha), NGO worker (Hassan), verifier (Clara).

The same test environment was used for consistency across all evaluation dimensions.

8.3 Usability Evaluation

8.3.1 Methodology

Usability was assessed through heuristic evaluation and user walkthroughs. Nielsen's 10 usability heuristics (Peffers et al., 2007) were applied alongside the Digital Dignity rubric (Pulse, 2021). Tasks included:

- Upload and redact a document
- Share the redacted version via WebID
- Revoke access to a shared document
- Inspect consent logs

8.3.2 Quantitative Findings

- **Task success rate**: 92%
- **Mean time to complete workflow (upload → share)**: 3.2 minutes
- **Error rate (e.g., sharing incorrect file)**: 8%
- **Satisfaction score (1–5 scale)**: 3.4 average

8.3.3 Qualitative Findings

Participants valued:

- The "revoke" feature for control over shared files
- Visual indicators of access status
- Offline readiness and file previews

Pain points included:

- Confusion over WebID vs email sharing
- Need for clearer labels on metadata logs
- Lack of language localization

8.3.4 Summary

The usability evaluation confirmed that the interface was largely intuitive and fit for purpose, with room for improvement in labeling and internationalization.

8.4 Security Evaluation

8.4.1 Threat Model: STRIDE

Security was evaluated using the STRIDE framework (Corp., 2022), which analyzes threats across:

- **Spoofing**
- **Tampering**
- **Repudiation**
- **Information Disclosure**
- **Denial of Service**
- **Elevation of Privilege**

8.4.2 Key Measures Implemented

- WebID-OIDC login with session tokens
- AES-256 encryption for original documents
- PLASMA-based access logs
- SHA-256 hash validation on document retrieval
- ACL-based access control
- Revocation logs and key invalidation

8.4.3 Penetration Testing Results

Simulated attacks included:

- URL tampering ($\rightarrow$ blocked by access control)
- Replay attacks ($\rightarrow$ blocked by session expiry)
- Metadata scraping ($\rightarrow$ RDF graphs require access grant)

No critical vulnerabilities were identified. All attack vectors were neutralized through existing controls.

8.4.4 Summary

The system meets a strong baseline for information security under realistic threat models. However, biometric spoofing and MFA are out of scope for this study.

8.5 Regulatory Compliance Evaluation

8.5.1 GDPR and DPDPB Requirements

The system was tested against the following legal principles:

- **Lawfulness, fairness, transparency**
- **Purpose limitation**
- **Data minimization**
- **Accuracy and retention control**
- **Rights of access, portability, and erasure**

8.5.2 Compliance Mechanisms

- Consent logs modeled in RDF via PLASMA
- Users can download all data and metadata
- Consent is revocable at any time
- Original files never exposed without consent
- No third-party analytics or tracking
- Hosting compliance metadata (jurisdiction, backup policy)

8.5.3 Audit Trail and Exportability

All metadata is SHACL-valid and machine-readable. A JSON-LD export of all interactions can be generated and submitted to third-party auditors.

8.5.4 Summary

The system fulfills core requirements of both GDPR (Union, 2016) and India's DPDPB (of India, 2023). Transparency and control are embedded into system architecture, not merely UI design.

8.6 Humanitarian Alignment

8.6.1 Ethical Principles Followed

- **Do No Harm**: Redaction and consent tools prevent oversharing
- **User Autonomy**: Users control sharing scope and can revoke it
- **Data Sovereignty**: Users retain full ownership of their Pods
- **Decentralization**: No NGO or state entity can monopolize identity

8.6.2 Alignment with UNHCR and ICRC Guidelines

The system implements:

- Participatory onboarding via NGO-assisted flows
- Transparent logging of access requests
- Rights-aligned language in the consent interface

8.7 Limitations of Evaluation Strategy

- Evaluations were conducted with proxy users, not actual refugees
- Penetration testing did not include biometric attacks
- Compliance logs assume trusted Pod providers — no legal enforcement simulated
- Real-world scaling not evaluated (tested with < 100 documents)

8.8 Conclusion

This chapter presented a multidimensional evaluation of the system's usability, security, legal compliance, and ethical alignment. Evaluation results affirm that:

- The system is usable by low-literacy users after brief orientation
- It withstands common security threats with robust cryptography and auditability
- It meets GDPR and DPDPB compliance through verifiable semantic metadata
- It operationalizes humanitarian data principles through consent, redaction, and decentralization

These findings collectively demonstrate that the system is not only technically viable but ethically deployable in humanitarian contexts. Its open-source, modular architecture and standards-based metadata ensure transparency, extensibility, and replicability for future field pilots or integrations with NGO identity ecosystems.

9 Discussion and Conclusion

9.1 Comparative Interpretation of Empirical Findings

This chapter integrates the findings in Chapters 5 to 8, providing a comparative analysis of how each of the sub-research questions contributes to the goal of designing an interoperable, secure, ethical, and field-viable identity system for forcibly displaced persons. While each chapter addressed separate dimensions—Solid's potential for decentralization (Chapter 5), usability and hosting concerns (Chapter 6), architectural enablers (Chapter 7), and real-world effectiveness (Chapter 8)—collectively, these findings encapsulate a shared vision and illuminate critical trade-offs and system-level synergies.

9.1.1 The Role of Solid Pods in Decentralized Document Ownership (Chapter 5)

Chapter 5 validated that Solid Pods enable a fundamentally different approach to identity management by creating an ecosystem of data that is centered on, and controlled by, the individual. Unlike centralized refugee registration platforms or NGO-controlled identity databases—where all documents and associated metadata are stored in institutional silos—Solid's Pod-based architecture ensures that documents are stored in spaces owned by the user, independent of any single organization, device, or geographic location. This structural shift removes the single point of control and failure inherent in centralized systems. By storing redacted and encrypted identity documents directly within user-owned Pods, the architecture actively mitigates risks such as document loss, unauthorized access, and indefinite institutional retention.

The "difference" also extends to how documents and their metadata are managed. In proprietary database schemas, metadata is often locked within the issuing organization's system, making interoperability and auditability difficult. By contrast, metadata in the prototype was modeled in RDF using established vocabularies (e.g., *schema.org*, DCAT, PLASMA), allowing any authorized application to interpret, verify, and trace document usage events in a standards-compliant way. This decentralized semantic model was not merely a technical abstraction—it operationalized user autonomy. Refugees could upload, redact, share, revoke, and audit access to their documents without surrendering control to an NGO, application developer, or backend provider.

The humanitarian implications are substantial. Under this model, the refugee shifts from being a passive "beneficiary" or "data subject" to an active *data steward*. Control over identity is no longer granted as a privilege by external institutions but is established as a technical default. This re-balancing of control, enabled by Solid's separation of data from applications and its fine-grained access controls, underpins the ethical foundation of the entire system.

9.1.2 User Interface and Hosting Considerations in Field Usability (Chapter 6)

Chapter 6 translated the ethical ideals of Chapter 5 into deployable interfaces and deployment strategies. It shows that decentralization alone isn't sufficient without an accessible, reliable, and responsive interface that illustrates the agency's promise at the architectural level.

Creating and testing a mobile-first, offline-capable PWA showed that with careful consideration of the user interface, and Solid's abstractions as WebID, grants of access, and RDF logs could be made accessible. The translation was not straightforward, however. Usability testing concluded that terms like "WebID" and "grant scope" confused people—even smartphone app-savvy people. This shows the tension between interface accessibility and protocol fidelity: the more a UI is made to look like the Solid spec, the less accessible it will be; the more that it's abstracted, the more likely to be abused or oversimplified.

Similarly, the choice of platform (Inrupt ESS vs CSS vs LocalPod) demonstrated that different deployment contexts require different levels of control, stability, and transparency. In high-connectivity environments, Inrupt offered a smoother experience, but in field conditions, hybrid or self-hosted servers were more reliable. This went towards confirming the architecture design

described in Chapter 7: one that decouples storage logic from user workflows and allows heterogeneous deployment without shattering user experience.

9.1.3 Architectural Design as a Bridge Between Ethics and Operability (Chapter 7)

Chapter 7 addressed the internal mechanisms of system architecture and how ethical decision-making must not only be based in policy but also in code and data flow. Layered architecture—in presentation, logic, metadata, storage, and access control—was a primary strength in reconciling ethical intent with functional implementation.

Selective disclosure, Chapter 5 definition, was supported by using redaction and consent-aware RDF metadata. Chapter 7 outlined how this reasoning was enforced not only within the UI but also at the RDF and file system level. Isolation of the redacted and encrypted forms, and permanent reference of both to a common semantic graph, allowed for both pragmatic usability and theoretical auditability.

Mobile-first constraints—discussed in Chapter 6—were natively supported by offline queuing, IndexedDB caching, and async sync protocols. Most importantly, these were achieved without undermining access control guarantees. After reconnection, the system replayed cached permissions in a way that retained revocation correctness and access logs. This close offline integration is an important addition to the state of Solid-based applications today, where offline situations are not typically supported.

Decentralized access control in the guise of Solid ACLs, WebIDs, and PLASMA became a technical enabler as well as an ethical standard. Having immediate representation of all permissions, revocations, and consent logs semantically (via vocabularies like FOAF, PROV-O, and PLASMA) ensured that data flows were always understandable, auditable, and user-controllable. This bridged the compliance-practice gap, which was explored further in Chapter 8.

9.1.4 Evaluating Effectiveness in Real-World Humanitarian Contexts (Chapter 8)

Chapter 8 capped the system design with a challenge by subjecting it to the pragmatic conditions under which it would be operating. With a multi-dimensional test and using a strategy—combining usability heuristics, STRIDE threat modeling, and GDPR/DPDPB compliance checklists—a tested for both functionality as well as its effectiveness in real-life scenarios.

The usability testing supported the results of Chapter 6: simplifying and localizing the UI enabled the majority of users to perform basic actions (upload, redact, share, revoke) within a five-minute limit. Sharing errors were rare and largely caused by ambiguity in access terminologies. The need for role-based UI flows and context tooltips is thus demonstrated.

STRIDE-based security testing revealed that the system was robust against common attack surfaces. Data leakage, spoofing, and privilege escalation were three challenges that were met by token-based access, encrypted channel for documents, and ACL checks. Though biometric

spoofing and device exploitation were out of scope, results guaranteed the architecture's tamper and unauthorized access resistance.

Regulatory compliance, as discussed in detail, was not an afterthought but one of the design axes. Consent logs, revocation records, access metadata, and hosting jurisdiction were all expressed as RDF and SHACL-validated. This renders compliance from a legal checkbox to an auditable, machine-verifyable characteristic of the data itself. This is orders of magnitude better in terms of traceability and accountability than static consent forms or UI-only confirmation dialogues.

Most strikingly, Chapter 8 tested the system for ethical preparedness. The system was not only tested on if it worked, but if it honored the dignity, autonomy, and rights of its users. This is an atypical metric for technical testing, and that it was included speaks to the interdisciplinary rigor of this research.

9.1.5 Synthesis Across Sub-RQs

Taken together, Chapters 5–8 form a very integrated sequence. Chapter 5 established the ethical and functional rationale for using Solid. Chapter 6 evaluated the usability and infrastructural impact of such an action. Chapter 7 presented the architectural realization of these plans in software modules. Chapter 8 assured the overall efficiency of the system under humanitarian conditions.

Cross-cutting themes that emerge include:

- **The interdependence of ethics and architecture**: Ethical goals like user agency are hollow without technical enforcement through access control, semantic logs, and revocable consent.
- **The need for modular, context-aware deployment**: A one-size-fits-all architecture is infeasible for a system expected to work in refugee camps, border crossings, urban shelters, and humanitarian offices.
- **The challenge of making Solid accessible**: While the protocol enables powerful decentralization, its terminology and mental models must be translated carefully to avoid user confusion and misuse.
- **The role of semantic metadata as a glue**: RDF and vocabularies like PLASMA not only enable interoperability, but encode legal and ethical states into the data itself.

Every chapter not only answers its sub-research question but also cross-checks other people's assumptions and constraints. For example, Chapter 7 architecture would not be an issue if Chapter 6 usability did not pass, and ethical goals in Chapter 5 would be irrelevant without Chapter 8 compliance testing. Consistency in such findings boosts internal consistency and theoretical tightness of the study as a whole.

9.2 Strength, Validity, and Limitations of Findings

While the empirical results presented in Chapters 5 through 8 demonstrate the feasibility and promise of a Solid-based refugee identity system, a rigorous analysis of their strength

and validity is essential to contextualize their broader implications. This section evaluates the reliability of the findings, the degree to which they can be generalized, and the key limitations that define their scope.

9.2.1 Internal Validity: Are the Results Credible Within the Study Context?

Internal validity refers to the confidence with which causal relationships or inferences can be drawn within the boundaries of the study's design and implementation.

- Controlled Testing Environment

  Usability, security, and compliance testing was conducted in controlled settings with pre-determined personas, same hardware (Pixel 3a, iPhone SE, Linux Chromebook), and simulated network conditions. Control enabled variable isolation and reproducibility. Functional modules were tested in isolation and combined into end-to-end workflows to enhance test coverage and traceability.

- Artifact Integrity

  The software artifact was built as modular, versioned codebase with reproducible environments (CI pipelines, Docker). RDF metadata were SHACL-validated and logs were auto-timestamped and auto-signed. These design practices reduce measurement error and ensure observed outcomes are caused by actual system behavior and not due to implementation defects or test variance.

- Test Repeatability

  All three test protocols (usability task flow, STRIDE threat simulation, SHACL audit compliance) were documented and reproducible. The task walkthroughs and persona profiles add assurance that other independent evaluators would be able to replicate the same results using the same configuration.

Conclusion on Internal Validity Overall, the study demonstrates high internal validity. The artifacts behaved consistently under a variety of controlled tasks, and the results can be causally attributed to the system's design. However, the human-centric components would benefit from future testing in live humanitarian settings.

9.2.2 External Validity: Can the Findings Be Generalized?

External validity concerns whether and how the results observed in this study can be generalized beyond the specific implementation context.

- Technological Portability

  The platform is independent of platforms, and it utilizes open standards—Solid, RDF, schema.org, PLASMA, SHACL. This implies the system architecture, metadata model, and redaction/encryption logic can be deployed on an array of hosting environments, device types, and jurisdictions. The technical architecture can then be generalized to humanitarian deployments.

- Diverse Hosting Environments

With experiments involving Inrupt ESS, CSS, and LocalPod configurations, the study demonstrated that critical functions (upload, share, revoke, access logs) remain resilient even under varying infrastructural conditions. This contributes to the confidence that the system could function within varying environments—ranging from refugee camps to city-based humanitarian offices.

- Cultural and Linguistic Diversity

  The current test was given in English and on Western-region-set hardware. This limits generalizability to linguistic, cultural, and educational settings. Although the UI was designed with translation hooks and minimal reliance on text, usage in the field would require to be implemented in a range of languages and literacy situations.

- Institutional Ecosystem Fit

  Human adoption among humanitarian actors is in its infancy. The architecture assumes that institutions (NGOs, border officials) can access and understand RDF metadata, process WebIDs, and apply consent-based workflows. Without institutional readiness, technically correct deployments may falter or be misunderstood.

Conclusion on External Validity The system is technically portable and architecturally sound across platforms, but social and institutional generalizability remains partial. Broader testing is required in multilingual, multi-actor, and culturally varied contexts to validate the system's robustness and acceptability at scale.

9.2.3 Strength of Contributions

- Technical Innovation The project brings with it a fully functional Solid-based system with offline-first support, double-file redaction logic, semantic consent modelling, and audit-ready RDF logs. These are not present or are exceptional in earlier Solid systems for humanitarian work. The system can also do decentralized, selective disclosure—something not too many systems are able to in a user-controllable way.

- Methodological Rigor The design procedure took the Design Science Research (DSR) paradigm (Peffers et al., 2007, Hevner et al., 2004) with iterative prototyping, multi-layered evaluation, and documentation of constraints. The integration of DSR with legal, ethical, and human-centered testing represents a methodological strength that spans technical and humanitarian research traditions.

- Ethical Grounding

  The system incorporates ethics into architecture, not as a list of outside-in rules, but by using them where they belong naturally—inside the architecture. Consent, revocation, and reducing data are used not just in UI flows, but at the semantic data layer. This shifts the field's knowledge of how to design privacy-preserving systems past checklists of compliance.

- Reproducibility and Extensibility

  All the code is module-based and documented, vocabularies, and test conditions. This makes it easy to adapt for other humanitarian NGOs, government agencies, or research groups who want to implement similar systems.

9.2.4 Limitations and Risks

- Simulated Users vs Real Refugees

  The largest constraint is the simulated users. While proxy personas reveal insight, they have no limitations in real life—emotional tension, device unreliability, digital dread—that may occur to refugees. Only a field trial can uncover these nuances.

- Limited Scale Testing

  Small-scale evaluation was conducted on small collections of documents (fewer than 100 documents per Pod). National-scale performance (e.g., multi-million document national refugee registries) is not tested.

- Legal Ambiguity of RDF Consent

  Semantic consent modeling is novel, but few jurisdictions formally implement RDF-encoded consent logs. Legal systems may necessitate traditional signatures or logs outside the Solid context. Harmonization by law in the future will be needed.

- Pod Governance and Key Recovery

  Solid currently lacks a standard for lost recovery of WebID or management of Pods in the case of death, coercion, or state capture. This adds risk: users can lose access forever or be denied portability as a result of lost credentials or inaccessibility of hosts.

Conclusion on Limitations The system is best viewed as a validated prototype with strong internal coherence and technical novelty. Its limitations—especially in legal formalism, institutional fit, and lived user testing—mark the boundary between academic feasibility and real-world deployment.

9.3 Conclusion: Answering the Overall Research Question

Overall Research Question

How can Solid Pods be used to design an interoperable, ethical, and technically resilient identity system for refugees?

9.3.1 Restating the Problem Context

The identity management problems of the refugees are multi-dimensional in nature: loss or confiscation of documents, non-interoperability across and between institutions and borders, primitive levels of digital literacy, ethical hazards of centralized control of data, and little infrastructure. The traditional digital identity systems—state registries or commercial platforms—are typically unsuitable for the humanitarian environment since they rest upon centralized control, rigid schemas, and opaque access processes.

As an answer, this thesis posed the question of whether Solid Pods—user-controlled, decentralized data stores based on Web standards—might be employed to create not only an operational, robust system but an ethical, privacy-respecting, and interoperable one. Four subquestions of research were covered in Chapters 5–8, each aiming at a specific face of the challenge.

9.3.2 Synthesis of Findings

- **Chapter 5** demonstrated how Solid Pods enable document preservation, sharing on the basis of consent, and decentralization. Users remain in charge even across shifting geographies and institutional settings. Redaction features and RDF-modeled metadata enabled fine-grained consent and auditability for supporting ethical data stewardship.
- **Chapter 6** determined that with a mobile-first approach and adaptive Pod hosting (ESS, CSS, LocalPod), Solid-based systems can be field-deployable and accessible. While there were some concepts (e.g., WebID) that ran counter to UX, role-based views and in-place help provided for learnability and trust.
- **Chapter 7** laid out a modular, layer-based architecture for offline-first access, selective disclosure, and semantic access control. Provenance and consent were not only UI features but were embedded into the data layer via vocabularies like PLASMA and DCAT.
- **Chapter 8** put the system to STRIDE security analysis, GDPR/DPDPB, and humanitarian principles (e.g., digital dignity) testing. The system was tamper-resistant, was private, and made consent logs machine-verifiable. While user trials are outstanding, proxy testing shows high readiness.

Together, these results affirm that Solid Pods can serve as the foundation for a secure, interoperable, user-governed refugee identity system—provided that system is supported by appropriate UI design, institutional training, semantic tooling, and legal/policy frameworks.

9.3.3 Strength of the Conclusion

Validity: Conclusion valid on the basis of high internal validity—modules performed consistently under testing conditions with reliable behavior and high semantic integrity. External validity is constrained: platform generalizability high but cultural and institutional generalizability need field trials.

Triangulation: Results are triangulated along a number of dimensions (usability, security, compliance), personas (refugee, NGO, verifier), and platforms (browser, mobile, local server), increasing confidence in findings.

Limitations: The key limitations is in-field deployment, limited integration with real humanitarian workflows, and the emerging maturity of the Solid ecosystem. These are significant, but do not invalidate the overall technical or ethical conclusions.

Practicality: The system is not a hypothetical design. It is an up-and-running, offline-capable PWA with deployed Pods, RDF metadata, and redaction/encryption logic. Its architecture and codebase can be reused or extended by NGOs, researchers, or governments wanting to deploy ethical identity systems.

9.3.4 Final Answer to the Research Question

Solid Pods can be used to design a refugee identity system that is interoperable, ethical, and technically resilient—provided the system is built using a modular architecture that supports selective disclosure, mobile-first access, and decentralized

semantic access control.

By decoupling data ownership from institutions, representing consent semantically, and embedding auditability into the data model, such a system can uphold user agency, enhance document security, and ensure long-term accessibility—even in fragile, multilingual, and institutionally complex settings.

9.3.5 Concluding Reflection

This thesis offers a rare synthesis of technical architecture, legal compliance, user interface design, and ethical grounding (of the Red Cross, 2020; Pulse, 2021; Union, 2016; Werbrouck, Pauwels, Beetz, Verborgh, and Mannens, 2024). It illustrates not just what is possible with Solid, but how to build it—step by step—in the service of vulnerable populations (Esteves and Pandit, 2023; for Refugees, 2022; Solid Project, 2023). It calls for a future where digital identity is not a source of exclusion or surveillance, but a tool for empowerment, mobility, and dignity (for Refugees, 2022; Pulse, 2021).

The road to large-scale adoption is long. But the work here provides a reproducible, standards-compliant, and ethically-informed blueprint for future research, pilots, and policy innovation (Schema.org, 2024; Sporny et al., 2019; Wilkinson et al., 2016).

References

Brickley, D., & Miller, L. (2024). Foaf vocabulary specification [Friend of a Friend project]. http://xmlns.com/foaf/spec/

Broeke, A. (2024). *Assessing migrant safety at eu external borders: Navigating the intersection of ai, policy, and human security* [B.S. thesis]. University of Twente.

Celuchova Bosanska, D., Huptych, M., & Lhotská, L. (2022). Decentralized ehrs in the semantic web for better health data management. In *Phealth 2022* (pp. 157–162). IOS Press.

Corp., M. (2022). The stride threat model. https://docs.microsoft.com/en-us/security/compass/stride-threat-modeling

Domingue, J., Third, A., & Ramachandran, M. (2019). The fair trade framework for assessing decentralised data solutions. *Companion Proceedings of The 2019 World Wide Web Conference*, 866–882.

Elrick, L. (2021). Finding the balance between security and human rights in the eu border security ecosystem. *European Journal of Law and Technology*, *12*(1), 1–42.

Esteves, B., & Pandit, H. J. (2023). Using patterns to manage governance of solid apps.

for Refugees, U. N. H. C. (2022). Unhcr data protection policy. https://www.unhcr.org/data-protection-policy

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105.

kanth Mandru, S. (n.d.). Data protection regulations and cross-border data transfer: Addressing the challenges of data protection in international data transfer.

Lang, I. G. (2024). Security-centric approach in the use of digital.

Li, Y., et al. (2021). Cross-border data transfer regulation: A comparative study of china and europe.

McGregor, L., & Molnar, P. (2023). Digital border governance: A human rights based approach.

of India, G. (2023). Digital personal data protection bill, 2023. https://prsindia.org/billtrack/the-digital-personal-data-protection-bill-2023

of the Red Cross, I. C. (2020). Professional standards for protection work. https://www.icrc.org/en/publication/0999-professional-standards-protection-work

Okoth, P. K. (2023). Security challenges in civil registration: Safeguarding vital information in an evolving landscape. *World Journal of Advanced Research and Reviews*, *19*(1), 1051–1071.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77.

Pulse, U. G. (2021). Guidelines for digital dignity in data systems. https://www.unglobalpulse.org/policy/guidelines-for-digital-dignity/

Rahimi, M. (2024). *A comprehensive analysis of privacy and data protection in conflict-affected areas: Revising human rights and humanitarian law to address the challenges of surveillance technologies* [Master's thesis].

Schema.org. (2024). Schema.org vocabulary [Accessed July 2025]. https://schema.org/

Shekar, K., Tripathi, A., Birla, B., & VAIDYA, E. (2023). Principle-based framework towards cross-border data transfers. *Available at SSRN 4527681*.

Solid Project. (2023). Solid technical specifications. https://solidproject.org/TR

Sporny, M., Longley, D., & Chadwick, D. (2019). Verifiable credentials data model 1.0. https://www.w3.org/TR/vc-data-model/

Third, A., & Domingue, J. (2023). Plasma: A vocabulary for provenance, licensing, access and security metadata audit. https://w3id.org/plasma/

Union, E. (2016). General data protection regulation (gdpr). https://eur-lex.europa.eu/eli/reg/2016/679/oj

Vavoula, N. (2024). Tr-ai-nsforming migration, asylum and border management in the eu: The roles of the ai act, interoperable large-scale it systems and eu migration agencies. *Interoperable Large-Scale it Systems and EU Migration Agencies*.

W3C. (2023a). Data catalog vocabulary (dcat) [W3C Recommendation]. https://www.w3.org/TR/vocab-dcat/

W3C. (2023b). Shapes constraint language (shacl) [W3C Recommendation]. https://www.w3.org/TR/shacl/

Werbrouck, J., Pauwels, P., Beetz, J., & Mannens, E. (2024). *The web as a common data environment: Management of federated multi-models* [Doctoral dissertation, Department of Electronics and Information Systems Faculty of Engineering and ...].

Werbrouck, J., Pauwels, P., Beetz, J., Verborgh, R., & Mannens, E. (2024). Consolid: A federated ecosystem for heterogeneous multi-stakeholder projects. *Semantic Web*, *15*(2), 429–460.

Werbrouck, J., Schulz, O., Oraskari, J., Mannens, E., Pauwels, P., & Beetz, J. (2023). A generic framework for federated cdes applied to issue management. *Advanced Engineering Informatics*, *58*, 102136.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018.

Ziyi, X. (2022). International law protection of cross-border transmission of personal information based on cloud computing and big data. *Mobile Information Systems*, *2022*(1), 9672693.