Artifact

Eduardo Oliveira

April 11, 2025

1 Open Science Platform

1.1 Overview

Traditional centralized research repositories often exhibit data silos, limited verifiability, and susceptibility to manipulation, impeding the openness and reliability of scientific practices. The decentralized model introduced in this work is designed to mitigate these challenges by enabling efficient data sharing, fostering collaboration, and enhancing the validation of research outputs, thereby strengthening reproducibility and transparency.

This chapter details the design and implementation of the Open Science Platform, a decentralized system that integrates blockchain, IPFS, and smart contracts to improve research reproducibility. By leveraging immutable records and decentralized storage, the platform ensures transparent and verifiable research artifact management. Additionally, extended services are incorporated to facilitate file indexing, metadata extraction, and search functionality. The proposed platform aligns with Open Science principles by providing verifiable and persistently traceable access to research artifacts. Figure 1 presents the high level building blocks of the platform.

1.2 Technology Stack

The Open Science Platform is built on a hybrid architecture that strategically integrates decentralized and centralized components to balance security, traceability, and efficiency in data management. Decentralized technologies, such as blockchain and IPFS, ensure data integrity and tamper resistance, while centralized components facilitate indexing, search, and user interactions. Figure 2 presents a high-level breakdown of the platform's core building blocks.

1.3 Core Services

The core services of the Open Science Platform provide the fundamental infrastructure for secure and verifiable research artifact management.

- **Hyperledger Iroha v1 Blockchain:** Acts as the immutable ledger for managing user and project accounts, recording transactions, and enforcing

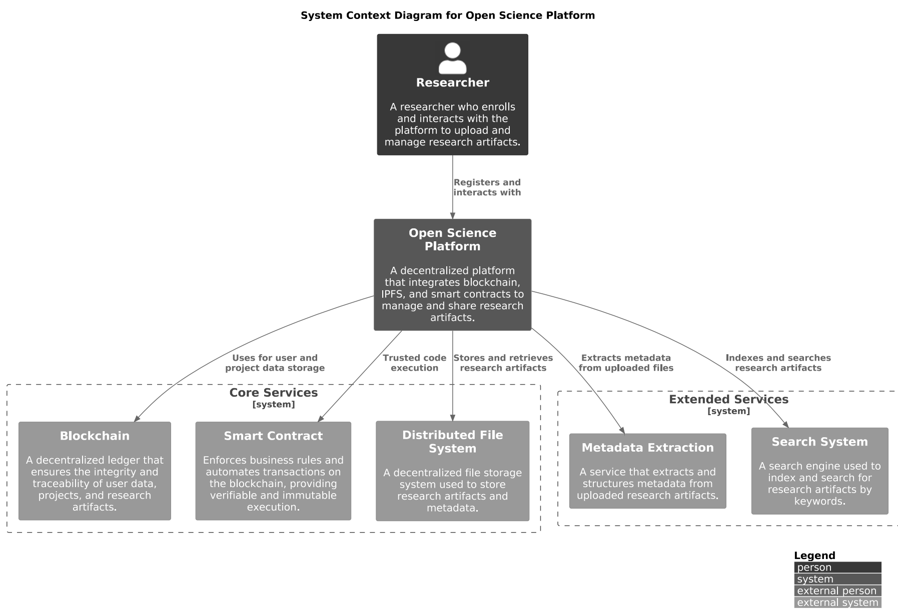


Figure 1: System context diagram for the Open Science Platform

business rules via smart contracts to ensure secure and transparent data exchange.

- **InterPlanetary File System (IPFS):** Provides decentralized, tamper-proof storage for research artifacts and metadata, ensuring persistent and verifiable access to shared data.

1.4 Extended Services

The extended services enhance the platform's features by improving file and metadata processing.

- **Apache Tika:** Extracts metadata from uploaded files, enhancing artifact organization and searchability.

- **Whoosh:** Facilitates efficient indexing and keyword-based search for stored artifacts.

1.5 User Interface, integration and execution

- **Jupyter Notebooks (Python):** Powers the front-end interface, facilitating the automation and display of the execution steps. Blockchain interactions are managed via the Iroha v1 Python library, while communication with the IPFS network is handled through the HTTPS client library.

1.6 System Components and Interactions in the Open Science Platform

The Open Science Platform consists of multiple interconnected components, each serving a distinct role in ensuring secure, verifiable, and reproducible research data management. The primary components include Jupyter Server, the blockchain Hyperledger Iroha v1 and the InterPlanetary File System (IPFS). Each of these elements are encapsulated within a Docker container to provide modularity, ease of deployment and reproducibility. The implementation level architecture is presented in figure 4, the network topology is depicted in figure 3

1.6.1 Jupyter Server

The Jupyter Server acts as the primary interface for users interacting with the platform. This component provides a Python kernel for the execution environment that integrates the Iroha v1 Library, the IPFS HTTPS client, Apache Tika for metadata handling, and the Woosh Indexer and Search system. It enables users to:

- Execute Python scripts to submit transactions and queries to the blockchain via smart contracts.
- Upload and retrieve files and metadata (JSON objects) stored in IPFS.
- Process and index research data using Apache Tika and Woosh for enhanced searchability.
- Access and visualize blockchain-stored metadata for Open Science applications.

1.6.2 Blockchain

The blockchain runs based on a Hyperledger Iroha v1 network and acts as a distributed ledger for recording transactions. It ensures immutability, transparency, and verifiability of stored research metadata. This component:

- Receives transactions from the Jupyter Server via a gRPC API.
- Stores metadata references, ensuring that uploaded research artifacts can be authenticated.
- Interacts with PostgreSQL for structured storage of blockchain metadata.
- Supports smart contracts through the integration of Hyperledger Burrow, which provides a modular blockchain client with a permissioned smart contract interpreter partially developed to the specification of the Ethereum Virtual Machine (EVM).

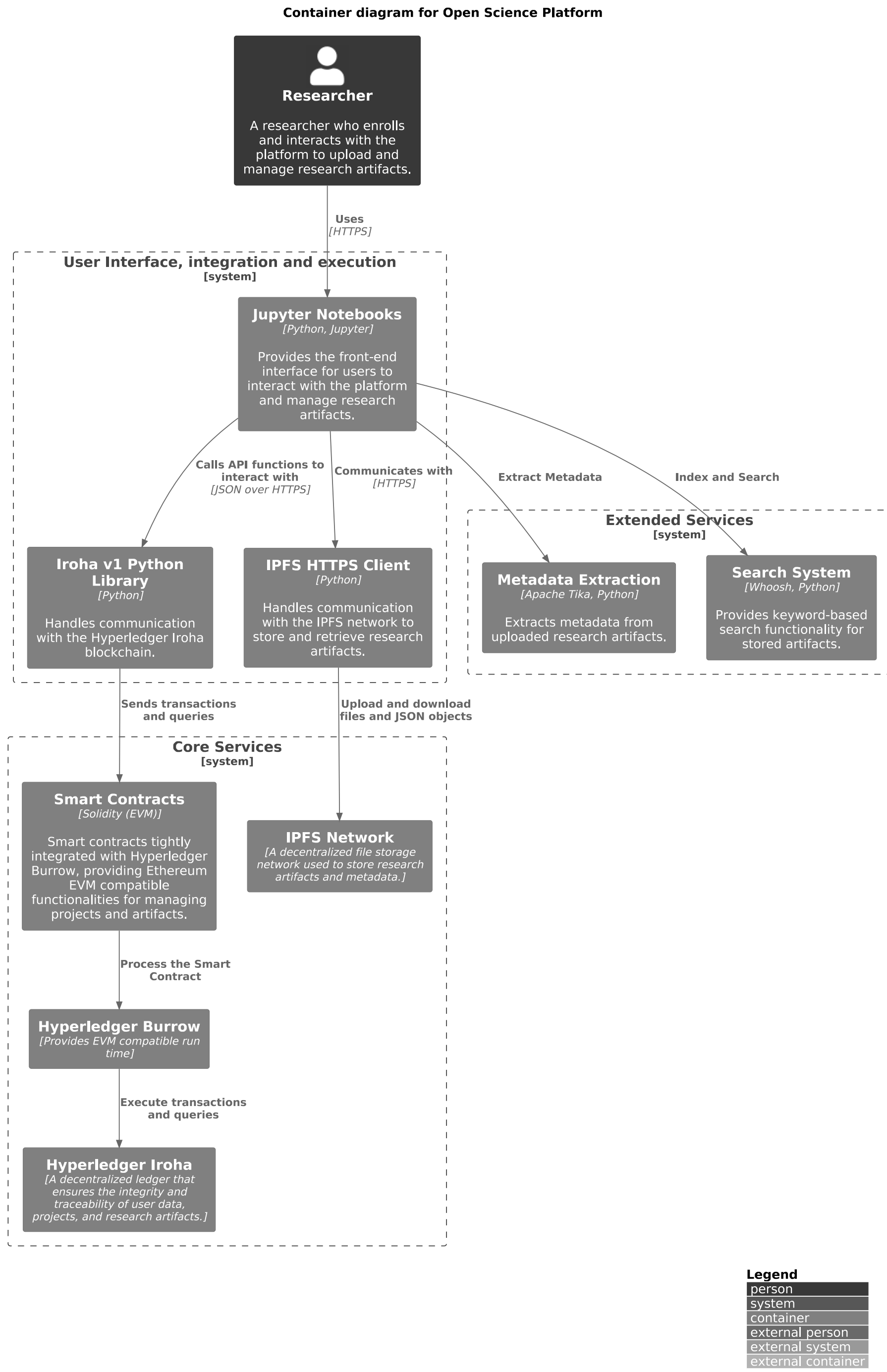


Figure 2: Container diagram for the Open Science Platform

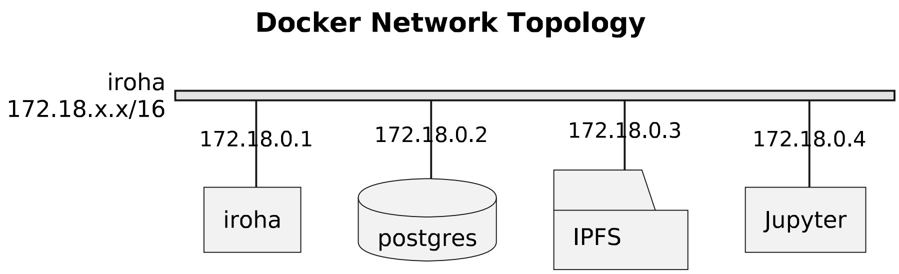


Figure 3: Docker network topology

1.6.3 Storage

The InterPlanetary File System (IPFS) is a decentralized storage solution that manages the research outputs. This component:

- Stores digital research artifacts in a content-addressed manner.
- Allows the Jupyter Server to upload and retrieve files via an HTTPS API.
- Ensures long-term availability of scientific data through distributed storage principles.

1.6.4 Relational Database (PostgreSQL)

The PostgreSQL database provides structured storage for blockchain-related data. It is used exclusively and managed by Iroha v1 to:

- Maintain an efficient and queryable record of transactions.
- Ensure that research metadata stored on the blockchain can be retrieved and verified.
- Support blockchain operations requiring fast access to structured data.

1.6.5 Component Interactions

The components interact in aseamless and decentralized manner:

1. **User Interaction**: The user submits transactions, uploads files, and queries research data through the Jupyter Server.
2. **Blockchain Transactions**: Jupyter Server sends and retrieves research metadata to the Iroha blockchain via gRPC API.
3. **Metadata Storage**: Iroha stores data in the PostgreSQL database for efficient retrieval.

4. **Decentralized Storage**: Research artifacts are stored in IPFS, with their unique file identifiers recorded on the blockchain.

5. **File Retrieval**: Users can retrieve files from IPFS using their content identifiers (CID), ensuring authenticity and reproducibility.

This architecture guarantees trustworthy and reproducible scientific research by leveraging blockchain for integrity, IPFS for decentralized storage, and Jupyter as an accessible research environment.

1.7 Platform Operations

The platform supports a set of core operations that regulate user interactions with projects and research artifacts.

1.8 User Enrollment and Project Registration

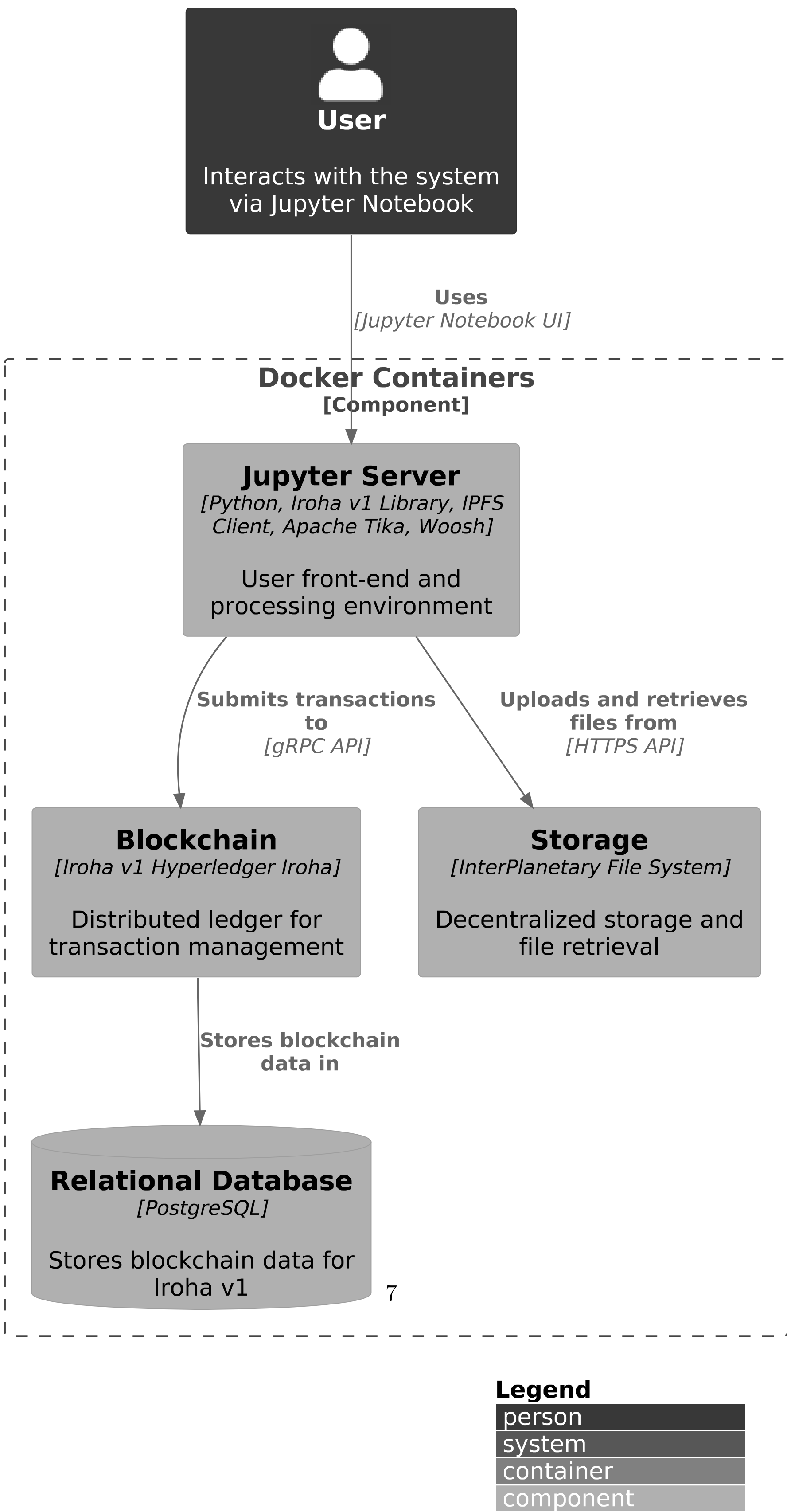
The Open Science Platform enables user enrollment and project registration, ensuring transparent and verifiable account management on the blockchain. Users self-enroll by providing cryptographic credentials and identity details, which are securely stored using a combination of blockchain attributes and decentralized storage through IPFS. Similarly, projects are registered with essential metadata, establishing a distinct blockchain account for each one. To maintain traceability and facilitate efficient project management, the system links user and project accounts bidirectionally, allowing for streamlined queries and provenance tracking. This process is depicted in Figure 5.

- **User Self-Enrollment** – A user self-enrolls on the platform by providing a private key that complies with the ED25519 or SHA-3 standards and identity information, including full name, institution, email, ORCID, and role. An account is created for the user in the blockchain. All data provided in the enrollment is structured in key/value pairs into a JSON object and uploaded to IPFS, with the corresponding Content Identifier (CID) recorded on the blockchain attributes of the user account.

- **Project Registration** – Users can register a project by specifying a descriptive name, an abstract, relevant keywords, start and end dates, funding agency, and location. Upon registration, a blockchain account is created. This data is structured in key/value pairs into a JSON object and uploaded to IPFS, with the corresponding Content Identifier (CID) recorded on the blockchain attributes of the project account.

- **User and Project Accounts Linkage** – Once both user and project accounts are created, the system updates their attributes to establish a bidirectional association. This ensures that querying a user account reveals linked project accounts, and vice versa, facilitating traceability and efficient project management.

Component Diagram for Open Science Platform



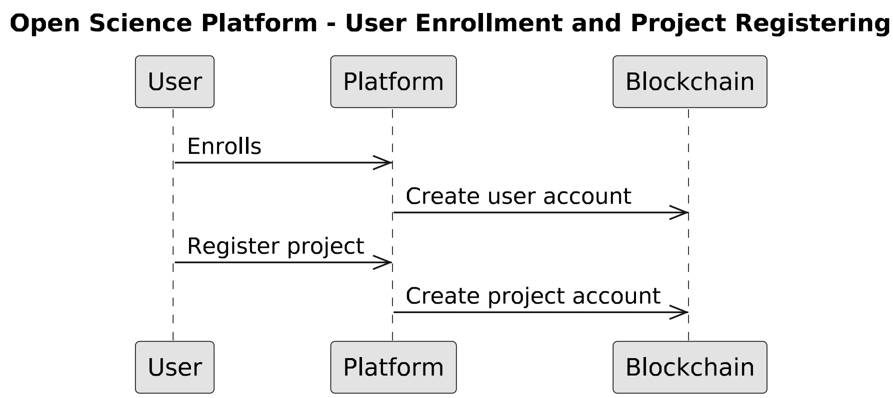


Figure 5: User enrollment and project registering for the Open Science Platform

1.9 Artifact Management

The Open Science Platform provides a structured approach to managing research artifacts, ensuring their integrity, traceability, and accessibility. Users can upload various types of research files, including papers, datasets, and images, which are securely stored in a decentralized manner using IPFS. Each uploaded file is assigned a unique Content Identifier (CID), which is recorded on the blockchain, creating a tamper-proof reference. To enhance discoverability, metadata is extracted, structured, and stored on IPFS, with its CID also registered on the blockchain. The system further supports indexing and full-text search capabilities, enabling efficient retrieval of research artifacts.

The file upload and metadata management workflow are illustrated in Figure 6. A user may upload research artifacts, such as papers, datasets, and images, which are stored on IPFS. Each file is assigned a unique CID, ensuring traceability and integrity, and this CID is recorded on the blockchain attributes of the corresponding project account. After upload, metadata is extracted, structured in key/value pairs, and uploaded to IPFS, with its CID also recorded on the blockchain to preserve provenance. To enhance searchability, the system indexes metadata, including full-text indexing for text-based files.

The platform also enables users to search for research artifacts using keyword-based queries. As depicted in Figure 7, the search engine looks up keywords in the indexed metadata and returns relevant results. Each result includes metadata details such as descriptions, subjects, and authorship, allowing users to identify relevant artifacts efficiently.

Once a file has been located, the platform performs a validation step to ensure its integrity and authenticity. The CID stored on IPFS is compared against the CID recorded on the blockchain. If they match, the file is considered valid; otherwise, the system flags it as potentially tampered with or corrupted. This validation mechanism safeguards research artifacts against unauthorized modifications. The file validation and retrieval process is depicted in figure 8.

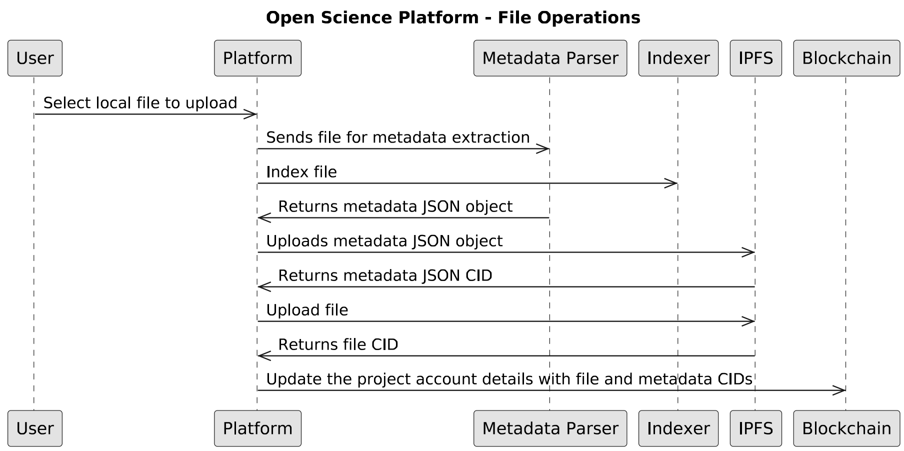


Figure 6: File operations diagram for the Open Science Platform

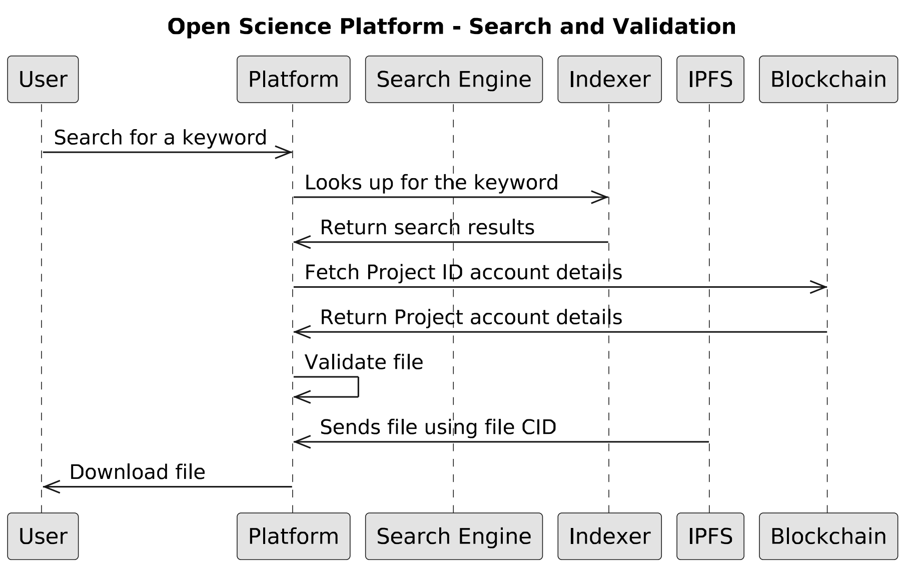


Figure 7: Keyword search, file validation, and download

A validated file can then be retrieved and downloaded from IPFS to the user's local system for further use.

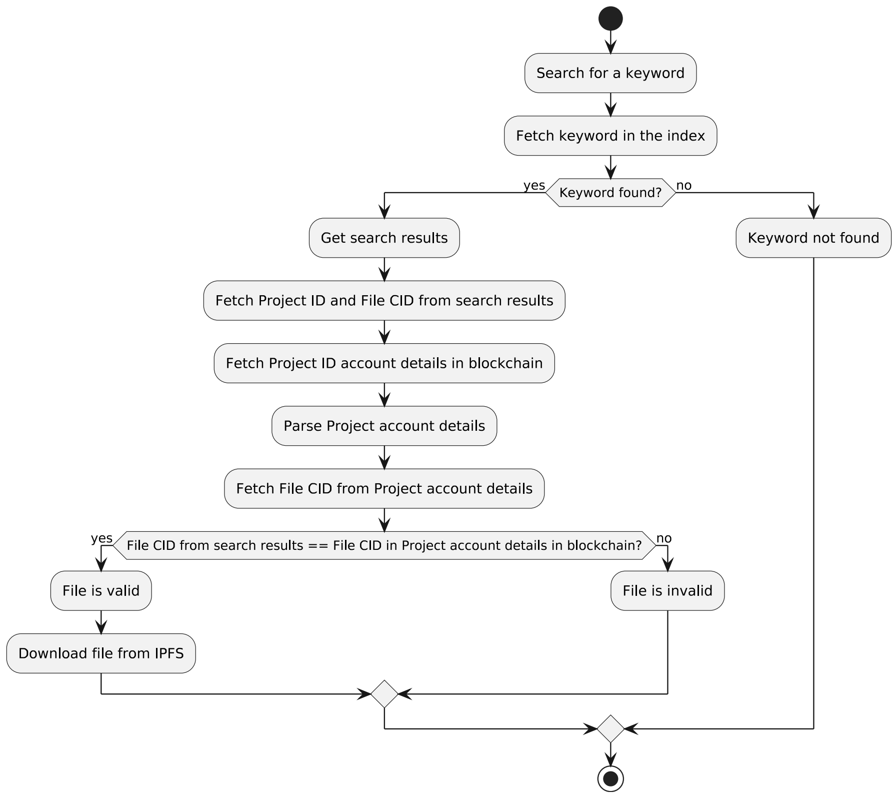


Figure 8: File validation and download

1.10 Data Model for the Open Science Platform

The entity-relationship model for the Open Science Platform defines the logical structure of users and research projects, capturing the associations between these entities. The primary entities in this model are `User` and `Project`, which are connected through an linked relationship, figure 9 presents the model.

1.11 User Entity

The `User` entity represents an individual interacting with the platform. Each user is uniquely identified by an account ID and has attributes that describe personal and institutional information. The attributes of the `User` entity are listed in Table 1.

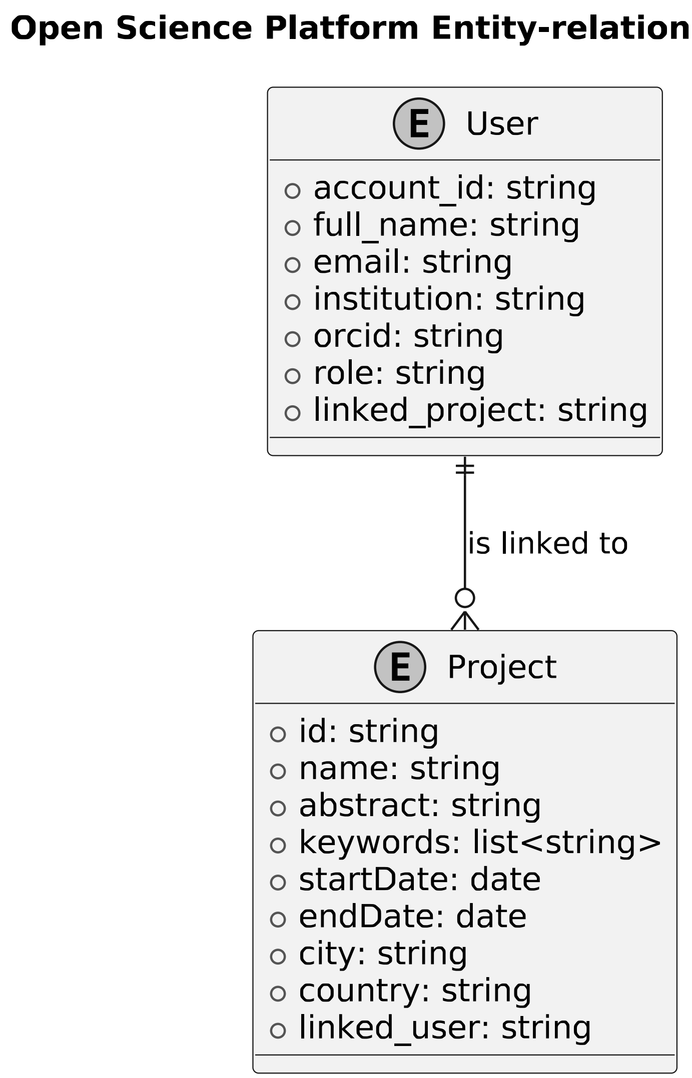


Figure 9: Entity-relationship model for the Open Science Platform

Table 1: User Entity Attributes

Attribute	Description
account_id	A unique identifier assigned to the user.
full_name	The complete name of the user.
email	The email address used for communication.
institution	The organization to which the user is affiliated.
orcid	The Open Researcher and Contributor ID.
role	The role of the user within the research project.
linked_project	The research project the user is assigned to.

1.12 Project Entity

The `Project` entity represents a research project registered in the platform. It contains essential metadata to describe the project and facilitate discovery and collaboration. The attributes of the `Project` entity are listed in Table 2.

Table 2: Project Entity Attributes

Attribute	Description
project_id	A unique identifier assigned to the project.
name	The official name of the project.
abstract	A brief summary outlining the research objectives.
keywords	A list of relevant keywords associated with the project.
startDate	The date when the project officially begins.
endDate	The date when the project was concluded or is expected to conclude.
city	The city where the project is primarily conducted.
country	The country associated with the research project.
linked_user	The user linked to the project.

1.13 Linked relationship between user and project enttities

A `User` is linked to one or more `Project` entities, establishing a one-to-many relationship. This means that a single user can be responsible for multiple projects. This model ensures a structured representation of research projects and their associated users, supporting an organized approach to data management in the Open Science Platform.

1.14 Data Model for Hyperledger Iroha v1

The entity-relationship (ER) model of Hyperledger Iroha defines the core entities, attributes, and relationships that facilitate role-based access control, asset management, and multi-signature security. While Iroha v1 includes a broader set of entities, this research focuses solely on the account and domain related classes and attributes, as presented in figure 10

2 Core Entities and Their Attributes

2.1 Account Entity

The `account` entity represents a user or system account registered on the blockchain. Table 3 lists its attributes.

Table 3: Attributes of the `account` entity

Attribute	**Description**
`account_id`	Unique identifier of the account
`domain_id`	Links the account to a specific domain
`quorum`	Required number of signatories for multi-signature transactions
`data`	Stores additional metadata in JSON format

2.2 Domain Entity

The `domain` entity organizes accounts within logical boundaries. A `domain` can contain multiple `accounts`, as illustrated in Table 4.

Table 4: Attributes of the `domain` entity

Attribute	**Description**
`domain_id`	Unique identifier for the domain
`default_role`	Default role assigned to accounts created within the domain

This ER model follows Hyperledger Iroha's permissioned blockchain structure. It ensures fine-grained access control, multi-signature security, and domain-based account management.

2.3 Relationship Between the Open Science Platform ER Model and the Iroha v1 ER Model

The Open Science Platform ER model leverages the entity structure of the Iroha v1 ER model, particularly the `account` entity, to represent both the `User` and

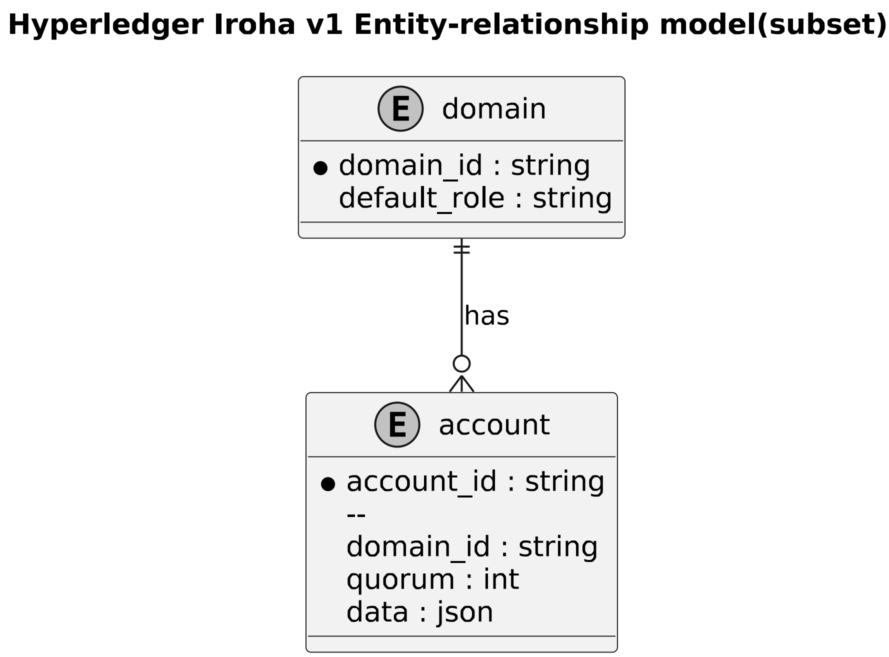


Figure 10: Subset of the Iroha v1 Entity-relationship model

`Project` entities. In this approach, instead of introducing separate entitties for users and projects, the `account` entity in the Iroha v1 ER model serves as a general-purpose representation, encapsulating all necessary attributes in a structured format.

The attributes specific to users and projects, which are not natively present in the Iroha v1 `account` entity, are stored as JSON objects within the `data` field of the `account` entity. This design provides a flexible and scalable means of extending the entity's attributes without modifying the core schema of the Iroha blockchain.

From a relational perspective, the `account` entity maintains its standard associations with roles, permissions, and assets as defined in the Iroha v1 ER model. This ensures that user accounts and project accounts can both participate in the blockchain's permissioning system, asset ownership model, and role-based access control without requiring modifications to the underlying structure.

By reusing the `account` entity, the Open Science Platform ER model ensures compatibility with Iroha's existing mechanisms for identity management, cryptographic signing, and permission delegation. Additionally, this approach aligns with the decentralized and immutable nature of blockchain, ensuring that both user and project entities benefit from the security and transparency features inherent to the Iroha v1 framework. Figure 11 provides a comparison between models and the rationale of use.

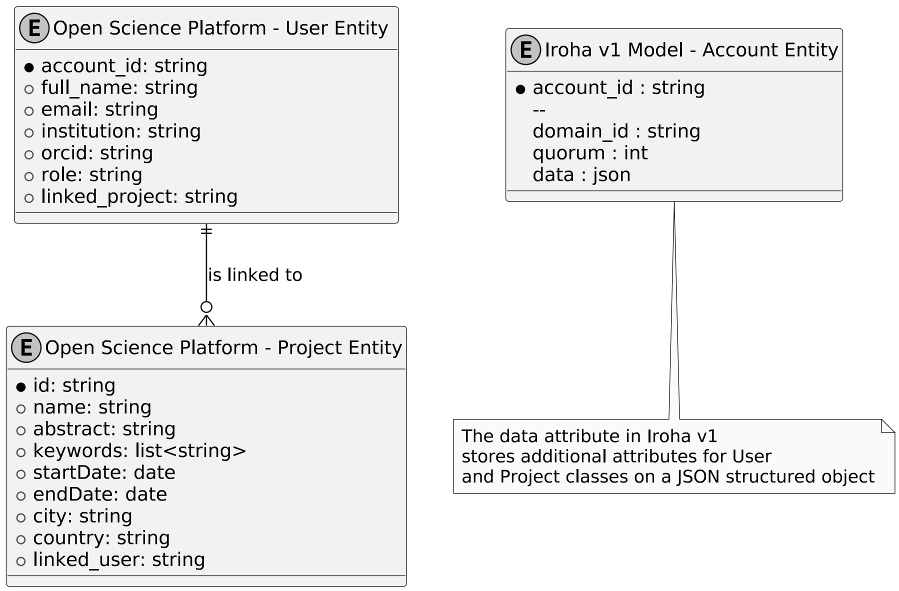


Figure 11: Comparison of the Entintiy-relationship models

2.4 The role of metadata and ontologies in the Open Science Platform

Metadata plays a crucial role in both the `Account` and `Project` classes within the Open Science platform. It is used to capture and represent essential information about the user and the research project, providing context and structure to their respective data. This metadata is stored as a JSON object [**?**], following established semantic web standards and leveraging ontologies to enhance data interoperability and accessibility.

2.5 Selected Ontologies

An ontology is a formal representation of knowledge as a set of concepts within a domain and the relationships between those concepts. In the context of the Open Science platform, ontologies help structure data in a way that promotes interoperability, consistency, and clarity. The use of ontologies such as `FOAF`, `Schema.org`, and `Dublin Core` ensures that data is standardized and can be easily shared and understood across different systems. These ontologies were chosen because of their widespread adoption, their ability to standardize data across different systems, and their support for rich, machine-readable representations.

As shown in table 5, by aligning with these ontologies, the platform ensures that its metadata is compatible with other Open Science initiatives and services, facilitating seamless integration and data exchange.

Ontology	Description
FOAF (Friend of a Friend)	A vocabulary used to describe people, their activities, and their relationships to other people and objects. It is used to describe the `User` entity, including attributes like name, email, and organization.
Schema.org	A collaborative initiative that provides a structured vocabulary for data markup on the web. It is used for describing both `User` and `Project` metadata, ensuring compatibility with web standards and promoting data discoverability.
Dublin Core (DC)	A metadata standard used for describing a wide range of resources, for describing the abstract, keywords, and other descriptive elements of the `Project` entity.

Table 5: Ontologies used in the Open Science Platform

2.6 User Metadata

The metadata for the `Account` class describes the attributes associated with a user on the platform. This metadata is structured using multiple ontologies, primarily `FOAF` (Friend of a Friend) and `Schema.org`, to provide detailed and interoperable information about the user. The key attributes in the `Account` metadata include the user's name, email, organizational affiliation, unique identifier (ORCID), role, public key, and linked project.

Attribute	Description
`foaf:name`	The name of the user.
`foaf:mbox`	The email address of the user.
`foaf:organization`	The organization the user is affiliated with, described as an instance of the `foaf:Organization` class.
`schema:identifier`	A unique identifier for the user, such as an ORCID identifier.
`foaf:holdsAccount`	The user's account details, including their role and public key.
`schema:linked_project`	The project associated with the user.

Table 6: Account Metadata Attributes

As presented in Table 6, this structured metadata helps ensure the user information is standardized and interoperable across different systems and platforms.

2.7 The use of JSON-LD for metadata representation

JSON for Linked Data (JSON-LD) is a lightweight Linked Data format designed to structure and interconnect data on the web using standard JSON. It extends JSON by incorporating semantic web principles, making data more discoverable, reusable, and machine-readable. JSON-LD achieves this by including a `@context` element, which maps terms to well-defined ontologies, and a `@graph` element, which structures entities and their relationships in a linked data format.

A key advantage of JSON-LD is its compatibility with existing JSON-based systems while enabling seamless integration with the semantic web. By leveraging vocabularies such as Schema.org and Dublin Core, JSON-LD ensures interoperability across diverse platforms and datasets. This makes it particularly useful for Open Science applications, where structured metadata enhances research reproducibility and data sharing.

In the context of the Open Science platform, JSON-LD is used to encode metadata for users and research projects, ensuring alignment with widely accepted ontologies. The structured representation enables automatic indexing, metadata enrichment, and semantic search capabilities, facilitating better knowledge discovery and integration within the scientific community.

2.8 The `User` Metadata JSON-LD object

The user metadata is structured using two primary ontologies: Friend of a Friend (FOAF) and Schema.org.

The FOAF ontology is used to describe personal and organizational attributes of users within the platform. It provides well-defined properties such as `foaf:name` for the user's full name, `foaf:mbox` for email addresses, and `foaf:organization` for institutional affiliations. By leveraging FOAF, the platform ensures standardized representation of user identities and their associations, facilitating integration with other systems that utilize FOAF-based user profiles.

Schema.org complements FOAF by enriching the user metadata with structured properties that enhance discoverability and machine readability. The `schema:identifier` property, for instance, is used to store unique user identifiers such as ORCID, ensuring compatibility with global researcher identification systems. Additionally, `schema:roleName` captures the user's role within the platform (e.g., reviewer, publisher), while `schema:publicKey` stores cryptographic keys associated with the user's account. The `schema:linked_project` property establishes connections between users and their associated research projects, enabling efficient metadata retrieval and knowledge graph construction as exhibited in Figures 12 and 13, the JSON-LD structure represents the project metadata in the Open Science platform.

Figure 12: JSON-LD structure for user metadata in the Open Science platform

```
{
    "@context": {
        "schema": "http://schema.org/",
        "foaf": "http://xmlns.com/foaf/0.1/"
    },
    "@graph": [
        {
            "@type": "foaf:Person",
            "foaf:name": "Zealous Ptolemy",
            "foaf:mbox": "zealous_ptolemy@email.com",
            "foaf:organization": {
                "@type": "foaf:Organization",
                "foaf:name": "Ashkelon Academic College"
            },
            "schema:identifier": {
                "@type": "PropertyValue",
                "propertyID": "ORCID",
                "value": "6153-7096-0437-X"
            },
            "foaf:holdsAccount": {
                "schema:identifier": "zealous_ptolemy@test",
                "schema:roleName": "reviewer",
                "schema:publicKey": "ca4c00c0a43bbd2caf070ab780886906ebb70e2c3d975972ccab4
            },
            "schema:linked_project": "02226@test"
        }
    ]
}
```

By combining FOAF and Schema.org, the Open Science platform ensures that user metadata is both human-readable and machine-actionable, promoting seamless integration with external research infrastructures and fostering an interoperable ecosystem for Open Science.

2.9 The `Project` Metadata JSON-LD object

The metadata for the `Project` entity provides essential details about the research project hosted on the platform. Similar to the user metadata, the project metadata is structured using `Schema.org` and `Dublin Core` (`dc`) ontologies. This structure allows for a comprehensive description of the project, including its name, abstract, keywords, timeline, funding details, and location.

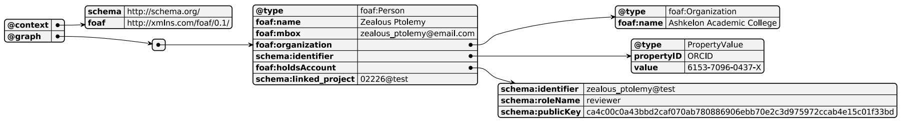


Figure 13: User metadata

Attribute	**Description**
`schema:name`	The name of the research project.
`dc:abstract`	A brief abstract describing the project's objectives and focus.
`schema:keywords`	Keywords associated with the project, such as "precision agriculture" and "global supply chains."
`schema:startDate`	The start date of the project.
`schema:endDate`	The end date of the project.
`schema:funding`	The funding organization for the project, described as an instance of the `schema:Organization` class.
`schema:location`	The physical location where the project is based, described as an instance of the `schema:Place` class.
`schema:metadataCID`	A unique identifier for the metadata of the project.
`schema:linked_user`	The user associated with the project.

Table 7: Project Metadata Attributes

The following JSON structure describes the metadata for a `Project` in the Open Science platform as shown in Figures 14 and 15.

This metadata not only captures the essential details of the project but also ensures that these details are linked to the user's profile, making it easier to track the relationship between users and their associated research efforts.

2.10 General Metadata Handling Workflow

The Open Science platform follows a general approach to metadata handling, ensuring that it is properly formatted, stored, and made immutable through blockchain integration. The process begins with processing the relevant metadata, which may pertain to a user, project, or file. This metadata is then formatted according to the JSON-LD standard, ensuring semantic interoperability and alignment with established ontologies.

Once formatted, the JSON-LD object is sent to the InterPlanetary File System (IPFS), a decentralized storage solution that provides content-addressable storage. Upon successful storage, IPFS generates a unique Content Identifier (CID) that serves as a reference to the stored metadata. This CID is then recorded on the blockchain by writing it into the account details associated with the entity. By anchoring the metadata CID on the blockchain, the platform ensures integrity, immutability, and transparency.

Finally, the blockchain transaction containing the CID serves as a provenance record, allowing stakeholders to verify and trace metadata modifications over time. The entire workflow guarantees that metadata remains both accessible and verifiable, promoting reproducibility and trust within the Open Science ecosystem.

Figure 16 illustrates the sequence of operations in the metadata handling process.

2.11 Blockchain Representation of Metadata

In the Open Science platform, metadata for users, projects, and files are stored on the blockchain. This ensures the integrity and provenance of the metadata while leveraging decentralized technologies. The following subsections describe the structure of blockchain representations for both user and project data, as well as the files associated with these projects.

2.12 User Account

The representation of a user's account on the blockchain contains the standard Iroha v1 attributes for the account entity, such as the unique account identifier, domain information, and quorum for consensus. Additionally, the `json_data` attribute references both the project to which the user is linked and the user's metadata CID (Content Identifier) stored on IPFS. This blockchain-based approach ensures that the user's information remains immutable and traceable, which is critical for maintaining the integrity of research data.

Figures 17 and 18presents the JSON structure for account details for a user in the Iroha v1 blockchain.

2.13 Project Account

The project account representation similarly uses a blockchain-based structure to store project-related metadata. Each project is identified by a unique account ID, along with the project's domain and quorum. The project metadata is linked to the user and includes important information about files associated with the project, including their CID references on IPFS. This ensures that the project data is linked to the user's account and that all files and metadata related to the project are securely stored on the blockchain for provenance tracking.

The JSON structure describes the account details for a user in the Open Science platform as shown in Figures 19 and 20.

File Representation

Within the project account, each file associated with the project is represented by a CID pair. The first CID refers to the file stored on IPFS, while the second CID references the metadata associated with that file. This structure ensures that the file's content and its metadata are both stored and tracked independently, but are still linked to the blockchain for integrity and provenance.

- The **first CID** (`QmTLZSqzPexwEdniZXLPN6fUfmEXX6MXS3b4QjKURgxc9y`) corresponds to the **file**.
- The **second CID** (`Qmchg7At5whR1T4xP8TwTMd8ntQqJXbbSicJRtGGaW1Z2P`) corresponds to the **metadata of the file**, ensuring that all relevant details are retrievable.

This structure allows for the efficient tracking and retrieval of research project data while maintaining provenance and integrity through blockchain storage.

2.14 Provenance in the Open Science Platform

The provenance system takes a two-fold approach, with both methods being native features of their respective systems. The first approach leverages Iroha v1's transaction logging capabilities, where each transaction is recorded with a hexadecimal hash and timestamp. This provides a reliable mechanism for tracking the evolution of account states over time. The hash acts as a snapshot, allowing for the retrieval of any past state of an account based on the corresponding transaction hash, as depicted in in Figure 21.

The second approach makes use of IPFS's native feature of Content Identifiers (CIDs) to track metadata associated with accounts, projects, and files. Each piece of metadata is linked to a unique CID, which allows for decentralized storage and immutability. A mismatch of the CID indicates that the metadata or file has been modified, ensuring the integrity of the information over time.

Together, these two approaches—transaction logging through Iroha v1's blockchain and metadata tracking through IPFS CIDs—provide a robust and transparent provenance system, ensuring both the transaction history and the integrity of metadata are verifiably recorded and traceable.

2.15 Conclusion

The Open Science Platform provides a comprehensive and robust framework for enhancing the reproducibility and transparency of scientific research. It leverages a modern technology stack comprising the Iroha v1 blockchain, InterPlanetary File System (IPFS), Jupyter Notebooks, Apache Tika, and Woosh. This stack ensures secure and efficient management of data and artifacts across the platform. The Iroha v1 blockchain, integrated with smart contracts and the Ethereum EVM compatible Hyperledger Burrow, guarantees the immutability and trustworthiness of all recorded actions, while IPFS enables decentralized storage of research data, ensuring scalability and resilience. Jupyter Notebooks serve as the primary front-end interface, providing an interactive environment for users to engage with research artifacts. Apache Tika facilitates the extraction and processing of metadata from various document types, while Woosh powers advanced search and indexing functionalities, improving data discoverability and retrieval.

Platform operations are streamlined and well-defined. User enrollment and project registration are handled seamlessly through the Iroha v1 blockchain, where both users and projects are registered as accounts. This facilitates the management of roles, permissions, and interactions within the platform, ensuring efficient tracking. Artifact management is integrated with the platform's metadata extraction system, allowing for the efficient storage and retrieval of research artifacts that are consistently linked to their respective provenance, reinforcing the auditability of scientific outputs. Additionally, the search and validation functionalities provide users with tools to find, and explore research data.

The data model, represented through an Entity-Relationship (ER) model, underpins the platform's data structure, offering a flexible and comprehensive approach to managing users, projects, metadata, and research artifacts. The role of metadata in the platform is crucial, as it is structured using well-established ontologies, including FOAF, Schema.org, and Dublin Core. These ontologies standardize the metadata representation, enabling interoperability and ensuring that all data is both machine-readable and discoverable. The integration of blockchain technology ensures that all metadata is transparently recorded, with the blockchain acting as a secure ledger for all metadata transactions, ensuring data integrity and facilitating trust in the platform. Provenance tracking is an essential aspect of the Open Science Platform, allowing for the tracing of data and results back to their origin, providing transparency, accountability, and enhancing the reproducibility of research.

Looking to the future, the Open Science Platform holds significant potential for further development. Ongoing efforts will focus on expanding the artifact management capabilities, integrating more advanced search functionalities, and enhancing the platform's ability to handle large-scale, multi-source research data. A key focus will be on improving interoperability with existing open science platforms and initiatives, ensuring seamless integration and data exchange across various systems. Additionally, the platform aims to incorporate blockchain-based incentive mechanisms to encourage active participation and adoption. By offering rewards and recognition for contributors and users, these mechanisms will help foster a collaborative and thriving research ecosystem, accelerating the platform's growth and encouraging wider use in the scientific community. The continued adoption of open standards and the enhancement of user experience will ensure the platform's scalability and relevance within the broader open science ecosystem.

Figure 14: JSON-LD structure for project metadata in the Open Science platform


```
{
    "@context": {
        "schema": "http://schema.org/",
        "dc": "http://purl.org/dc/terms/"
    },
    "@graph": [
        {
            "@type": "schema:ResearchProject",
            "schema:identifier": "02226@test",
            "schema:publicKey": "1c6b8d00c8382c93eb0dd3eeb24a20bfece56a28326bbaebb647cadaf
            "schema:description": {
                "@context": {
                    "schema": "http://schema.org/",
                    "dc": "http://purl.org/dc/terms/"
                },
                "@type": "schema:ResearchProject",
                "schema:name": "Assessing the Benefits of precision agriculture for global
                "dc:abstract": "This research focuses on the benefits and challenges posed
                "schema:keywords": [
                    "precision agriculture",
                    "global supply chains",
                    "disease prevention"
                ],
                "schema:startDate": "2023-12-18",
                "schema:endDate": "2027-01-02",
                "schema:funding": {
                    "@type": "schema:Organization",
                    "schema:name": "World Wildlife Fund"
                },
                "schema:location": {
                    "@type": "schema:Place",
                    "schema:name": "Los Angeles, California, USA"
                }
            },
            "schema:metadataCID": "Qmay4cDaxUaZaHoJKqzN69XkiX8wMx17aG4VMmwmkLcL1a",
            "schema:linked_user": "zealous_ptolemy@test"
        }
    ]
}
```

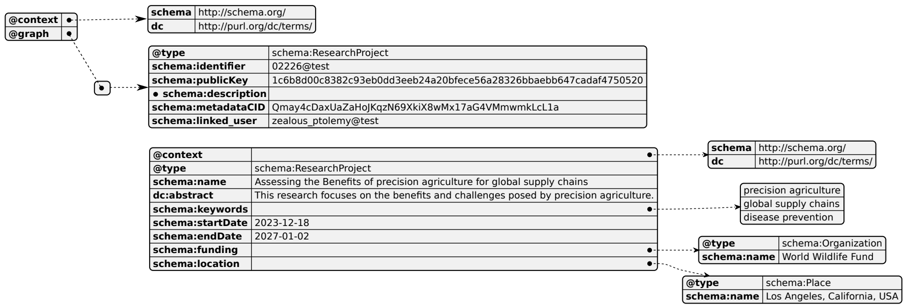


Figure 15: Project metadata

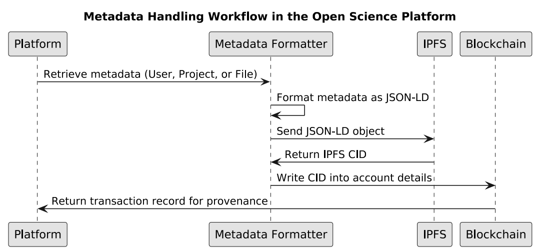


Figure 16: General workflow for metadata handling in the Open Science Platform

Figure 17: Blockchain Representation of User Account


```
{
    "account_id": "zealous_ptolemy@test",
    "domain_id": "test",
    "quorum": 1,
    "json_data": {
        "admin@test": {
            "linked_project": "02226@test",
            "account_metadata_cid": "QmT31fzDBNYAz1jAoAa7gQqSP7mDquv3fR8z1xLfxeHR5o"
        }
    }
}
```

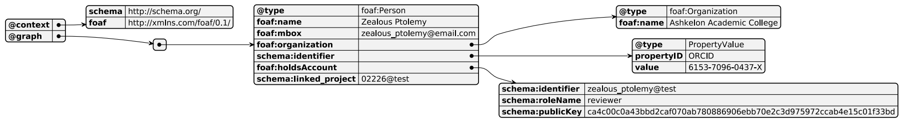



Figure 18: User blockchain representation

Figure 19: Blockchain Representation of Project Account


```
{
    "account_id": "02226@test",
    "domain_id": "test",
    "quorum": 1,
    "json_data": {
        "admin@test": {
            "file_1": [
                "QmTLZSqzPexwEdniZXLPN6fUfmEXX6MXS3b4QjKURgxc9y",
                "Qmchg7At5whR1T4xP8TwTMd8ntQqJXbbSicJRtGGaW1Z2P"
            ],
            "linked_user": "zealous_ptolemy@test",
            "account_metadata_cid": "Qmay4cDaxUaZaHoJKqzN69XkiX8wMx17aG4VMmwmkLcL1a"
        }
    }
}
```

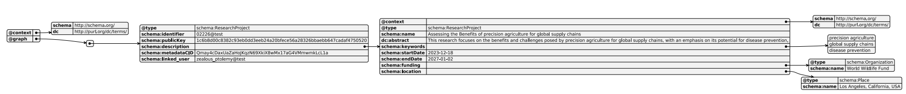



Figure 20: Project blockchain representation

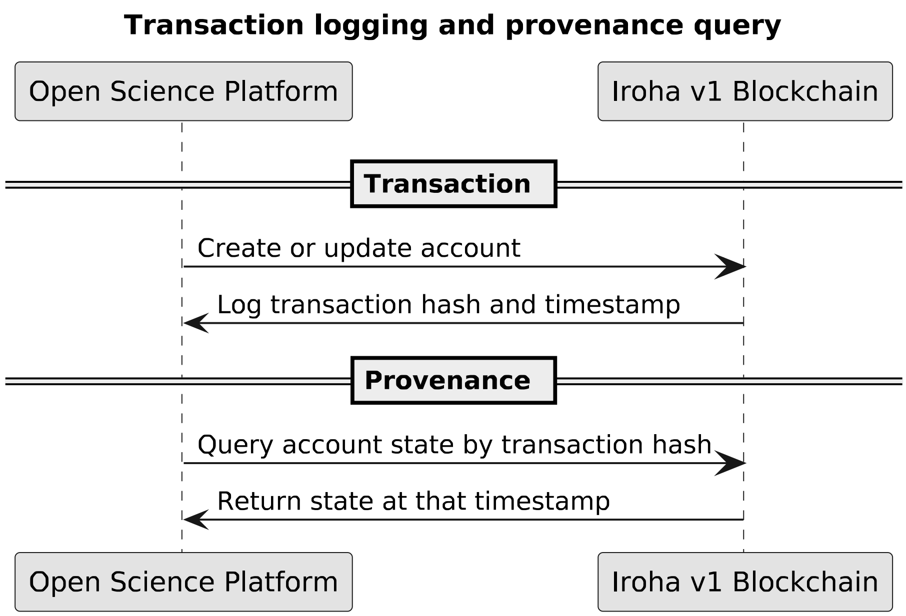


Figure 21: Transaction logging and provenance query