universidade de aveiro deti departamento de eletrónica, telecomunicações e informática

Project DiSA (Digital Signed Archive)

André Cardoso,
andremacardoso@ua.pt,
108269

Bruno Páscoa,
brunopascoa03@ua.pt,
107418

Maria Sardinha,
mariasardinha@ua.pt,
108756

Miguel Pinto,
miguel.silva48@ua.pt,
107449

Pedro Rei,
pedrorrei@ua.pt,
107463

Tiago Figueiredo,
tiago.a.figueiredo@ua.pt,
107263

Advisor Teachers,
José Vieira & André Zúquete

May 12, 2024

1 Abstract

In the digital age, the demand for efficient document submission faces persistent challenges due to storage limitations and file upload constraints. To face these obstacles, our project seeks to develop an innovative digital document archiving platform, dedicated to redefine document management with a constant commitment to authenticity.

Our initiative revolves around the development of a robust platform capable of accommodating digital documents, primarily in the versatile PDF format, while steadfastly preserving their integrity and authenticity. At the heart of our vision lies the concept of authenticity, encompassing the assurance that documents once archived, remain unaltered and trustworthy over time.

Key features include a streamlined document submission process, fortified by advanced digital signatures and blockchain technology to ensure authenticity verification. Moreover, our platform will provide the functionality to assign unique links to documents, simplifying referencing and verification.

Seamless integration with the Paperless platform will further enhance the resilience and accessibility of archived documents. To ensure user satisfaction, we prioritize the development of an intuitive user interface, enabling effortless document submission, retrieval, and authenticity verification.

Supported by the Library and STIC services at the University of Aveiro, our effort aims to revolutionize digital document management, empowering users with a secure, efficient, and user-friendly digital document archival solution, backed by authenticity guarantee.

Keywords: Digital Document Archiving, Document Authenticity, PDF Format, Digital Signatures, Blockchain Technology, Streamlined Submission Process, Unique Document Referencing, Paperless Integration, User-friendly Interface, Document Management Innovation

Contents

2 Introduction

TODO

2.1 Context

In today's digital landscape, organizations/users across diverse sectors are embracing digitization to streamline their operations and enhance efficiency. This paradigm shift towards digital processes has caused a large proliferation of digital documents, starting from reports and contracts to research papers and administrative records.

However, with the growing quantity of digital documents comes a set of challenges that organizations/users must address. Firstly, there may be the pressing issue of storage limitations, as the amount and size of digital documents grow, organizations/users face the daunting task of managing and storing those files securely and cost-effectively. Public institutions, in particular, encounter significant hurdles due to restricted storage capacities and stringent upload restrictions imposed by digital platforms.

Moreover, alongside the need for efficient storage solutions, there may be a crucial requirement to ensure the authenticity and integrity of digital documents. With the convenience of digital enhancing and manipulation, verifying the credibility of documents becomes increasingly complex. Users need robust mechanisms to authenticate the origin and integrity of digital documents, especially in contexts where document trustworthiness is paramount, such as legal agreements or academic publications.

Addressing those challenges demands innovative solutions that not only provide ample storage space for digital documents but also offer seamless document handling and authentication mechanisms. Organizations/users require a comprehensive digital document management platform that simplifies the storage, retrieval, and verification processes while upholding the authenticity and integrity of archived documents.

In response to those pressing needs, our project, DiSA (Digital Signed Archive), endeavors to develop a cutting-edge digital document archiving platform. By leveraging advanced technologies and user-centered design principles, our platform aims to revolutionize document management by offering scalable storage solutions, intuitive document handling features, and robust authentication mechanisms, thereby empowering organizations/users to manage their digital documents with confidence and efficiency.

2.2 Motivation

The motivation behind our project, "Digital Signed Archive" (DiSA), stems from the pressing need to address the challenges posed by the digitization of document management. As organizations/users increasingly transition towards digital processes, the volume of digital documents continues to grow exponentially. However, this surge in digital content brings forth several critical issues that need to be resolved.

Firstly, traditional document storage methods are no longer sufficient to handle the vast quantities of digital documents being generated. Organizations, especially public institutions, face constraints in terms of storage capacity and file upload limitations imposed by digital platforms. Consequently, there is a need for innovative solutions that provide scalable and cost-effective storage options for digital documents.

Secondly, ensuring the authenticity and integrity of digital documents has become a significant concern. With the ease of digital manipulation, verifying the credibility of documents has become increasingly challenging. It's imperative for organizations to have robust mechanisms in place to authenticate the origin and integrity of digital documents, particularly in sectors where document trustworthiness is paramount, such as legal and academic domains.

Moreover, the process of handling and managing digital documents needs to be streamlined and efficient. Users require intuitive platforms that simplify document submission, retrieval, and verification processes. Additionally, seamless integration with existing digital platforms enhances accessibility and usability, further improving the overall user experience.

By addressing these challenges, our project aims to revolutionize digital document management by developing a comprehensive and user-centered digital archive platform. We strive to provide organizations/users with the tools and functionalities needed to securely store, authenticate, and manage digital documents, thereby enhancing productivity, reliability, and trust in digital document workflows.

2.3 Goals

Our project aims to achieve a comprehensive set of goals that collectively redefine digital document management and archival practices:

1. Simplify Document Submission: Facilitate the process of sending digital documents, ensuring security and practicality. This involves developing an intuitive user interface that allows users to upload documents quickly and securely.

2. Guarantee Authenticity and Integrity: Ensure the authenticity and integrity of digital documents over time through the use of digital signatures and blockchain technology. This will prevent tampering and provide users with confidence in the authenticity of archived documents.

3. Scalable and Secure Storage: Eliminate limitations on the size and storage capacity of digital documents. Our platform will provide scalable storage solutions to accommodate documents of large sizes, ensuring long-term preservation and accessibility.

4. Seamless Access and Retrieval: Ensure easy access and retrieval of documents for users. Users should be able to retrieve documents they have submitted, while should also be able to access documents shared with them by others.

5. Enhanced Security and Privacy: Implement robust security measures and access control policies to ensure the security and privacy of archived documents. This includes exclusive sharing of selected documents and strict access control policies.

6. Intuitive User Interface: Develop an intuitive interface that caters to users of all ages and backgrounds, allowing them to perform tasks quickly and efficiently. User interface design will be user-centered, ensuring ease of use and accessibility.

7. Interoperability and Compatibility: Ensure interoperability with existing systems and standards to provide a seamless user experience. Integration with systems such as Paperless will enhance the overall functionality and interoperability of the platform.

8. Auditing and Version Control: Implement auditing and version control mechanisms to track access and changes to documents. This will provide transparency and accountability, allowing users to track document history and verify authenticity.

2.4 Expected Results

We anticipate the following outcomes from our project, reflecting the successful achievement of our interconnected goals:

1. Enhanced User Experience: Users will experience a streamlined and efficient document submission process, resulting in increased user satisfaction and engagement.

2. Improved Document Security: The implementation of digital signatures and blockchain technology will ensure the authenticity and integrity of archived documents, enhancing overall document security.

3. Scalable Storage Solutions: The platform will provide scalable storage solutions, allowing users to archive large volumes of data without constraints.

4. Easy Access and Retrieval: Users and organizations will have seamless access to archived documents, facilitating document retrieval and sharing.

5. Increased Privacy Protection: Strict access control policies will safeguard the privacy of archived documents, ensuring that only authorized users have access to sensitive information.

6. Enhanced Interoperability: Integration with existing systems and standards will enhance interoperability and compatibility, providing users with a seamless document management experience.

7. Transparent Document History: Auditing and version control mechanisms will provide transparency and accountability, allowing users to track document history and verify authenticity.

8. Long-term Document Preservation: Utilization of secure storage platforms like Paperless will ensure the long-term preservation of archived documents, guaranteeing their accessibility and integrity over time.

2.5 Document Structure

TODO

3 State of the Art

3.1 Introduction

In the realm of digital document management, users face persistent challenges in ensuring the authenticity, integrity, and long-term preservation of their digital assets. While various solutions exist to address specific aspects of document management, few integrate both the security of digital signatures and the immutability of blockchain technology into a comprehensive platform.

Our project, Digital Signed Archive (DiSA), aims to bridge this gap by developing an innovative digital document archiving platform that combines the security of digital signatures with the immutability of blockchain technology. By integrating these cutting-edge technologies, DiSA seeks to revolutionize document management, offering users a secure, efficient, and user-friendly solution for storing, authenticating, and preserving digital documents.

3.2 Related Work

In our exploration of existing solutions in the field of digital document management, we identified several notable platforms and technologies:

3.2.1 Document Archive Platforms

- **DSpace**: DSpace is a free open-source repository commonly used by academic institutions to manage and store various types of documents such as datasets, theses, images, and more.
- **Archivematica**: Archivematica is an open-source platform dedicated to the preservation of digital content. It provides comprehensive tools for managing and preserving digital archives, ensuring their long-term accessibility and integrity.
- **Preservica**: Preservica is a paid platform that supports multiple formats, including images and videos. It is utilized by institutions requiring long-term storage and preservation of their documents.
- **Paperless**: Paperless is a platform dedicated to secure and permanent document storage. It offers features for organizing and preserving documents, ensuring their accessibility and integrity over time.

3.2.2 Digital Signature Platforms

- **Blockcerts**: Blockcerts is an open standard that leverages blockchain technology to issue and verify digital certificates. However, it relies on the existence of an issuing institution.

- **Proof of Existence**: Proof of Existence is a platform that enables users to submit a document's hash onto the Bitcoin blockchain, providing proof of the document's existence at a specific point in time. It is available both as a standalone platform and as an npm package.

3.2.3 Other Notable Solutions

- **JSign**: JSign is a solution for digitally signing documents using blockchain technology. While it offers blockchain-based signatures, it is limited to a web-based service.
- **DocuSign**: DocuSign is a leading player in the eSignature market, focusing primarily on electronic signatures. However, it may lack integrated support for blockchain-based authentication.

While these solutions excel in specific areas of document management, none fully integrate both digital signatures and blockchain technology into a unified platform like DiSA aims to do. By combining the strengths of these technologies, DiSA seeks to provide organizations with a comprehensive solution for secure, authentic, and long-term document management.

3.3 Metrics

TODO

3.4 Technologies

TODO

3.4.1 Technology 1

TODO

3.4.2 Technology 2

TODO

3.4.3 Technology 3

TODO

3.4.4 Technology 4

TODO

3.5 Conclusions

TODO

4 Product Concept

TODO

4.1 System Requirements

TODO

4.1.1 Requirements Elicitation

TODO

4.1.2 Context

The system is designed to facilitate the management and preservation of digital documents for both individual users and organizations. It aims to provide a seamless experience for document submission, authentication, sharing, and retrieval while ensuring the integrity and long-term preservation of the documents.

4.1.3 Actors

- **Actor 1 - John Smith:** John is a teacher who wants to apply to work at the University of Aveiro. He already has his application files ready and needs to submit them through the service.
- **Actor 2 - Emily Davis:** Emily works in HR for the University of Aveiro. She frequently receives applications through links, which helps economize mailbox space. Therefore, she needs to access the service to retrieve each applicant's documents.

4.1.4 Use Cases

- **Use Case 1 - Document Submission**
  - **Actor**: John Smith (document submitter)
  - **Motivation**: John wants to store his application documents in a secure and trustworthy way.
  - **Preconditions**: John is authenticated in the system and owns a digital document ready for submission.
  - **Actions**:
    1. John selects the documents to be submitted.
    2. The system requests the digital signature of the documents.
    3. John digitally signs the documents.

    4. The system processes the documents and uploads them.
    5. The system generates a receipt linking original and processed documents (through hash) containing a unique link.
    6. The system stores the receipt in the blockchain and displays it to John.
    7. John is given the ability to share and/or save the unique link obtained (and its receipt).
  - **Post-conditions**: The documents are stored securely and authenticated with a unique unique link for access and validation.

- **Use Case 2 - Document Search**
  - **Actor**: Emily Davis
  - **Motivation**: Emily wants to get a document (a teacher's application).
  - **Preconditions**: Emily has a unique link (given by the system).
  - **Actions**:
    1. Emily accesses the system using the unique link.
    2. The system verifies the link's authenticity (if it is valid/exists) and Emily's permission to access the document.
    3. The system recovers the documents associated with the link given.
    4. The system provides the document(s) to Emily, allowing visualization/download.
  - **Post-conditions**: Emily has access to visualize/download the documents while being assured they are authentic and untampered with. (Notice: Sometimes the document cannot be accessed until a certain date)

- **Use Case 3 - Document Update**
  - **Actor**: John Smith
  - **Motivation**: John wants to update his previously submitted documents.
  - **Preconditions**: John is authenticated and has permission to update the document.
  - **Actions**:
    1. John selects the document to be updated.
    2. The system asks for the digital signature of the new version of the document.
    3. John signs the new version of the document.

    4. The system validates the digital signature and records the new version of the document on the blockchain.
    5. The system updates the unique unique link to point to the new version of the document.
    6. The system confirms the update and displays the updated unique link.
  - **Post-conditions**: The document is securely updated, and the unique link reflects the most recent version of the document.

- **Use Case 4 - Document Exporting and Sharing**
  - **Actor**: John Smith
  - **Motivation**: John wants to export or share his submitted documents.
  - **Preconditions**: John is authenticated and has permission to export or share the document.
  - **Actions**:
    1. John selects the document to be exported/shared.
    2. The system requests John's authentication to confirm permission.
    3. The system allows John to choose the exportation format of the document and the sharing method (DOI link, directed sharing).
  - **Post-conditions**: The document is exported/shared as requested by John.

- **Use Case 5 - Access Control**
  - **Actor**: John Smith
  - **Motivation**: John wants to control access to his submitted documents.
  - **Preconditions**: John is authenticated and has permission to configure the access control to the document.
  - **Actions**:
    1. John selects the document and defines its access level.
    2. The system requests John's authentication to confirm permission.
    3. The system updates the document's access configurations.
    4. The system confirms the update of the access configurations.
  - **Post-conditions**: The document has the access configurations updated as defined by John.

- **Use Case 6 - Document Access**
  - **Actor**: Emily Davis
  - **Motivation**: Emily wants to access and verify the authenticity of a document received through a unique link.
  - **Preconditions**: Emily has a unique link to access the document.
  - **Actions**:
    1. Emily inserts the unique link in the system.
    2. The system searches for the document corresponding to that unique link.
    3. The system verifies the document's authenticity.
    4. The system shows the document for access.
  - **Post-conditions**: Emily has access to the document and can confirm its authenticity.
- **Use Case 7 - Document Retrieval**
  - **Actor**: Emily Davis
  - **Motivation**: Emily wants to retrieve documents associated with a unique link.
  - **Preconditions**: Emily has a unique link (given by the system) associated with a specific document/set of documents and wants to obtain it/them.
  - **Actions**:
    1. Emily accesses the system using the unique link.
    2. The system verifies the link's authenticity and Emily's permission to access the document.
    3. The system recovers the document(s) associated with the link given.
    4. The system provides the document(s) to Emily, allowing visualization/download.
  - **Post-conditions**: Emily has access to visualize/download the document(s). If Emily does not have permissions, the system denies access and notifies her.
- **Use Case 8 - Change Tracking**
  - **Actor**: Emily Davis
  - **Motivation**: Emily wants to track the changes made to a document received through a unique link.
  - **Preconditions**: Emily has a unique link to a document and wants to track its changes.

- **Actions**:
  1. Emily inserts the unique link in the system.
  2. The system searches for the document's change history in the blockchain.
  3. The system displays the change history.
- **Post-conditions**: Emily has access to the document's alteration history, allowing her to track the changes over time.

4.1.5 Functional Requirements

- **Document Submission:**
  - The system must allow users to submit individual documents or sets of documents, preferably in PDF format. Documents with larger sizes than the standard (e.g., up to 100MB) should also be supported.
- **Authenticity Guarantee:**
  - The system must implement a mechanism to ensure the authenticity of submitted documents using digital signatures. This mechanism should prevent changes to documents after submission without altering their date and allow users to verify the authenticity of documents at any time.
- **Proof of Existence:**
  - The system must incorporate a mechanism to prove the existence of a document at a given time. This could be achieved through integration with a blockchain or similar technology.
- **Way to Share:**
  - The system should provide a way for users to associate a unique link with each document or set of documents. This unique link facilitates reference and verification of document authenticity by providing a permanent and unique identifier.
- **Persistence Archive:**
  - Integration with Paperless, an archive platform, is essential for the system to ensure the persistent archiving of documents. This integration should include support for the ingestion process, ensuring the preservation and accessibility of documents over time.
- **User Interface:**

    - The system should feature an intuitive and user-friendly interface that instills confidence in users. The user interface design should prioritize ease of use and accessibility for users of all levels of technical proficiency.

- **Receipt Generation:**

    - The system must automatically generate a receipt for each submitted document. This receipt should establish a link between the downloadable document and its submitted counterpart, providing users with a means to verify the authenticity of the downloaded document.

4.1.6 Non-Functional Requirements

- **Performance:**

    - The system must be capable of supporting a large volume of documents, including big documents, without experiencing degradation in performance. It should ensure fast response times for submission operations, authenticity verification, and document access.

- **Security:**

    - Robust security measures must be implemented to protect users' documents and information. This includes measures such as data encryption, strong authentication mechanisms, and intrusion detection systems to safeguard against unauthorized access and data breaches.

- **Reliability:**

    - The system must ensure high reliability and availability at all times. It should include redundancies and backups to prevent data loss in case of system failures or disruptions.

- **Usability:**

    - The user interface must be designed to be easy to use, allowing users to perform their tasks efficiently. It should provide intuitive and clear feedback to guide users through the document management process.

- **Compatibility:**

- The system must be compatible with different web browsers to ensure that users can access documents and perform operations from various environments. It should support popular browsers to maximize accessibility.

- **Maintenance and Documentation:**
  - The project should include comprehensive documentation to facilitate maintenance and future upgrades. This documentation should be well-explained and easily accessible, providing guidance on system operation, configuration, and troubleshooting.

4.1.7 Mockups

TODO

4.1.8 Brainstorming

TODO

4.1.9 Risks and Dependencies

TODO

4.2 System Architecture

TODO

4.2.1 Domain Model

TODO

4.2.2 Architecture Design

TODO

4.2.3 Deployment

TODO

5 Implementation

TODO

5.1 Presentation Layer

TODO

5.1.1 User Interface Design

TODO

5.1.2 UI Components and Layouts

TODO

5.1.3 Data Presentation and Visualization

TODO

5.1.4 Error Report and Feedback

TODO

5.1.5 Integration with the Business Layer

TODO

5.2 Authentication Layer

TODO

5.2.1 Authentication Layer 1

TODO

5.2.2 Authentication Layer 2

TODO

5.3 Business Layer

TODO

5.3.1 Business Layer 1

TODO

5.3.2 Business Layer 2

TODO

5.4 Database Layer

TODO

5.4.1 Database Layer 1

TODO

5.4.2 Database Layer 2

TODO

6 Project Management

TODO

6.1 Team Roles

TODO

6.2 Communication Plan

- **Repository:**
  - GitHub
- **Within Group:**
  - Meetings: At least one per week (more if needed) - Discord
  - Conversations: Everyday (almost) - WhatsApp group
- **With the Advisors:**
  - Meetings: 1 schedule per week (more if needed) - Zoom
  - Conversations: When necessary - Email by the group representative

6.3 Project Schedule

TODO

6.4 Methodology

TODO

7 Tests and Results

TODO

7.1 Static Code Analysis

TODO

7.2 Unit Testing

TODO

7.3 Usability Testing

TODO

7.3.1 Sample

TODO

7.3.2 Method

TODO

7.3.3 Results

TODO

7.4 Prototype Testing

TODO

7.4.1 Sample

TODO

7.4.2 Method

TODO

7.4.3 Results

TODO

8 Conclusions and Future Work

TODO

8.1 Summary

TODO

8.2 Main Results

TODO

8.3 Limitations

TODO

8.4 Conclusions

TODO

8.5 Future Work

TODO

9 References

TODO