Digital Identification: Creating a Document Management System

CS4991 Capstone Report, 2025

Akira Durham
Computer Science
The University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia USA
zup9su@virginia.edu

ABSTRACT

Innova8, a Northern Virginia tech startup, addressed inefficiencies in current document management systems, such as verification delays and human error, by creating a novel offline-capable application. To develop offline capability, I utilized PDFBox, an open-source PDF generation Java library, alongside various timestamping authorities, both of which enabled features such as document sealing, long-term validation, and document generation in the application. While this project was focused on the minimum viable product, development began with the core feature of verifying PDF documents and generating documents. Following the implementation of core features, development branched into accessibility focused areas such as a mobile app, adding barcodes for ease of use, as well as creating APIs for remote and application use. Results from the initial development include the completion of a minimum viable product to take to market, Fortune 500 company interest, and increased venture capital interest throughout the United States. Second phase development could involve tailoring features towards specific company specifications or use cases, improving the user interface, and integrating pilot test feedback into the core product.

1. INTRODUCTION

What if a missing timestamp, unverified signature, or an outdated system caused a document to be invalidated, delaying critical business operations or creating legal disputes for a family? In industries where accuracy and timing are important to operations such as healthcare, business, and law, document failures can have devastating consequences from a contract falling through to someone receiving the wrong medications. Despite these dangers, companies continue to use outdated document management systems that are vulnerable to human error and lack easily verifiable validation indicators. Without necessary checks in place digital document management and security is not ensured, leaving room for malicious actors and typos to propagate through the document management systems of today.

Innova8 aims to solve this problem through the development of a document management system that will address document inefficiencies and weak document security. By utilizing PDFBox for Java, an open-source PDF generation and editing library, this solution ensures long-term validation, online and offline confirmation ability, as well as easy validation of documents.

2. RELATED WORKS

During the initial development, I conducted research into previous document management systems to improve on the prior work as well as discover new features to incorporate into the design. Loesch and Theodori (2005) discussed how increases in paper documents would lead to the creation of a document management

system. They suggested that consolidating these documents into a central location would ensure reliability, consistency, and security in the management of these documents; however, the lack of security is a key factor that was overlooked, as people could alter the documents prior to uploading. Their idea for a digital document management system was innovative, especially for 2005. My project builds upon this idea of central storage for critical documents; however, my project addressed the issue of document security by incorporating hashing, long-term validation, and digital signatures to ensure a tamper-proof storage option.

From a more recent lens, some experts claim that electronic document management systems need to be specialized for the use case and that the three main considerations are: management, security, and control (Ismael & Okumus, 2017). While I incorporated these ideals into the product, Ismael and Okumus overlook the need for additional security to prevent tampering in sensitive documents and consideration for ways to securely share these documents. To combat these concerns, I incorporate a two-tier strategy of blockchain storage for document hashes and place e-seals on the documents to indicate when a document has been altered from its signed state.

Similarly, Ugale, et al., (2017) propose the use of an electronic document management system that incorporates security and management features; however, they focus closely on the security and accessibility aspects. These authors address key issues from previous papers by incorporating verification, retrieval, and access patterns into their design, accounting for the weak links presented in the previous two articles. Building on these ideas, I include several of these considerations into my document management system such as the access and verification steps. However, I addressed accessibility standards that were not present in their management system, allowing a wider range of people to make use of this product; and I included an offline verification component.

Previous work in the domain provided key insights into standard implementation considerations, as well as avenues for improvement. By combining these ideas, I was able to address common issues and build upon the current base of knowledge in document management systems.

3. PROJECT DESIGN

The electronic document management system builds upon the previous works, creating a security and document verifiable focused application. By addressing the previously identified issues for a secure document management system, I was able to create a novel product that is poised to further eliminate paper-trails, reduce human error, and increase document handling efficiency.

3.1 Requirements

The idea for this system began as an observation of current document workflow inefficiencies and insecurities. Diving head-first into the development phase without identifying weaknesses of current systems and power-user needs, the system would not fully address current downfalls. To discover these insights, I met with several potential clients, investigating what their current processes did well, as well as possibilities for improvements. Combined with my mentor's industry knowledge and hands-on experience with document management systems, I was able to generate guidelines and requirements for the system including steps for document generation, verification, and distribution. These requirements set the stage for the concentrated development of the system from the start.

3.2 System Development
By planning the development of this system with these goals in mind, I laid the groundwork for the core functionality, a PDF utility application that would provide features such as document issuance, verification, and retention using PDFBox. Development began with the goal of sealing a document to prevent tampering post-generation, providing peace of mind and security for the creator of the document. This core idea addressed concerns brought up from the previous works such as long-term validity and ensuring stored documents were not altered between the user uploading files and another user accessing those same files. After creating this core module, it became an integral part of the development cycle, providing direction as I continued to add additional features.

The next major feature in the development cycle was allowing users to generate and verify documents, consisting of a two-part process, generation followed by verification, ensuring that the generation encountered no errors, and that the system could verify its own documents. Considering the need to verify documents, I incorporated a time-stamping authority, DigiStamp, to facilitate long-term validation as verified from a trusted source and keep a consistent timing mechanism for printing on the documents, preventing delays from creating invalid documents.

Following the creation of these methods, I developed a series of APIs so that the team working on the front end and, eventually, users could pass all required information through the program securely and conduct thorough testing during the generation and web app processes. These API methods allowed the continuation of the project through mobile development as well as increasingly modular code, furthering ideas of encapsulation and maintainable code.

The overall development of this system was focused on addressing security concerns within electronic document management systems, as well as reducing human error in document generation. The system created accomplishes these tasks and builds on them via mobile and web applications, implementing a security focused document management system.

4. ANTICIPATED RESULTS
At the time of completion for my section of the project, immediate results include the completion of a minimum viable product containing the previously mentioned features, leading to venture capital interest. While this is a good result economically, technologically this system indicates a shift away from paper-based documents following the trend over the past 20 years in technological evolution within digital document management systems.

By implementing solutions to key security concerns in electronic document management systems, I address the current systems' weaknesses, allowing for further pilot test specific implementation or refinement from user feedback. Through user and organization feedback, the system would enter a testing development phase, where the feedback would be reviewed and integrated based on overall contribution to the core goals of the system and popularity amongst testers.

This testing phase would allow for more efficient implementation of impactful features and refine the system with real-world use-cases. Anticipated results include additional unit testing of existing security, feature development, and integration within existing document management systems, allowing for the eventual replacement of these older systems.

5. CONCLUSION

This project focused on addressing a lack of security within electronic document management systems. To increase security, I implemented key features such as long-term validation in document generation, secure document sharing, and creating tamper-proof documents. Through a secure document management system, sensitive documents can be handled with efficiency, prioritizing speed, accuracy and security combined with ease of use. By operating as a standalone product, this system can be implemented into current document systems easily, enabling the phase-out of older models using the APIs to migrate current documents and ensure document integrity during migration.

Through the development of this project, I learned how to work with security features within PDF documents, how to integrate with other developers across projects to create a cohesive product and how to thrive within a faster-paced tech startup environment.

6. FUTURE WORK

Moving forward, this project should be tested in a real-world environment including initial setup, deployment and extended use to ensure security claims, as well as get user-feedback. This testing could be accomplished through pilot tests because of venture capital interest or companies looking to upgrade their electronic record document management systems. Other steps to expand on the project could include further security developments such as implementing more secure API communications and streamlining the project package to reduce bloat and increase modularity.

REFERENCES

Ismael, A., & Okumus, I. (2017). Design and implementation of an electronic document management system. *Mehmet Akif Ersoy Üniversitesi Uygulamalı Bilimler Dergisi*, *1*(1), 9-17.

Loesch, J., & Theodori, J. (2005). Document management: a case study. *Journal of Facilities Management*, *3*(3), 273-283.

Ugale, M., Patil, S., & Musande, V. (2017, October). Document management system: A notion towards paperless office. In *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)* (pp. 217-224). IEEE.