KUKINI: THE CHALLENGES IN THE DESIGN, IMPLEMENTATION, AND EVALUATION OF A DIGITAL RECORDS TRANSFER TOOL FOR THE HAWAII STATE DIGITAL ARCHIVES

A THESIS SUBMITTED TO THE
GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAI‘I AT MĀNOA
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

INFORMATION AND COMPUTER SCIENCES

MAY 2014

By

Keone Y. Hiraide

Thesis Committee:

Philip Johnson, Chairperson
Luz M. Quiroga
Rich Gazan

To my family,

Leilani Y. Hiraide, Jay J. Hiraide, Lili S. Hiraide,

To Buma who may rest in peace,

Thank you for all your unprecedented support.

ACKNOWLEDGMENTS

I would like to thank Adam Jansen and Todd Blume for their guidance and advice during Kukini's development. I would also like to thank Dongie Agnir, Calvin Wong, Micah Takabayashi, and Todd Blume for reviewing Kukini's code as I made iterations. I feel that working on this project improved my skills as a programmer and a software developer.

ABSTRACT

At the Hawaii State Archives, there is a need to update their digital records preservation capabilities. Thus, they are currently in the process of implementing a records system which has been designed to store, protect, and preserve digital records. The types of digital records include medical records, annual reports, birth records, etc. This records system requires a Digital Records Transfer tool which must provide government agencies of Hawaii with the ability to transfer digital records to the Hawaii State Archives. Its transfer process must use secure and authenticated methods that document and ensure that the entirety of the files have been transferred uncorrupted. Kukini is a digital records transfer tool that has been designed, implemented, tested, and evaluated for use within an archival framework. This paper discusses the design, implementation, and evaluation of Kukini.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Over the last half century, there has been a rise in the use of various software technologies by organizations, institutions, and people throughout the world. These technologies allow information to be "born" digital and distributed electronically to a wide array of people through machines, the Internet, and other mediums. Organizations such as the Hawaii State Archives are in the process of updating their paper-centric processing/managing procedures to account for the complexities of handling digital information, in the form of "digital records". Before discussing the aforementioned complexities recognized by the Hawaii State Archives, I will briefly provide background information and discuss concepts surrounding the authenticity of digital records and their preservation. This will provide insight into the The Hawaii Digital Archive's need for the development of a Digital Records Transfer Tool such as Kukini, which is the main topic and premise of this paper.

1.1 Definition of Digital Records

What is a record? According to InterPARES 2 Terminology Database [3], a record is a, "document made or received in the course of a practical activity as an instrument or a by-product of such activity, and set aside for the action or reference." Jansen [22] adds that "Records have always possessed an innate value in that they serve as evidence of the activities in which they participate." In other words, although records are data, not all data are records. Records differ from data in that the main use of records is to document the past for the purpose of serving as evidence. Digital records can then be thought of as a subset of a record. Digital records differ from records solely on their medium; they are stored, handled, and transferred electronically.

Digital records are currently used in a variety of fields, organizations, and institutions. For example in the medical field, electronic medical records are used to document patient medical history, test results, and billing information. Health providers refer to the electronic health records of patients to learn about their medical health background in order to provide the best care and medical advice. Thus, it is important that patient medical records accurately capture up-to-date patient health history. In law enforcement, digital records are used in forensics, court systems, and in general practice as evidence in professional work. The validity and authenticity of these digital records are of major importance in the jurisdiction of criminal cases. In organizations such as the Hawaii State Archives, their main mission is to store and protect records and digital records of permanent value from loss, alteration, deterioration, and technological obsolescence so that they may be accessible in the future. The common characteristic that is pertinent to each of the examples mentioned is that digital records serve as important information whose purpose is to be referenced as evidence. Thus, because digital records are used as evidence, they must be authentic, reliable, and dependable. Park [29] states that, "In order to maintain this evidentiary capacity that is originally inherent in the records, those records must be demonstrably authentic, that is, the record must be intrinsically able to be proved that it is what it purports to be." The following section will discuss the complexities involved in the process of ensuring that digital records are authentic and valid with high integrity.

1.2 The Authenticity, Validity, and Integrity of Digital Records

As discussed in the previous section, it is important that digital records be authentic so that they may be used as relevant evidence. But how does one distinguish and identify a record as being authentic? According to Jansen [22], "The authenticity of a record is a factor of establishing the record identity and demonstrating the integrity of that record." He goes on to later state that, "The identity of a record is derived from the whole of its attributes that when taken together characterize and distinguish that record from others like it. Integrity is a reference to the degree that the record is complete and uncorrupted in all essential respects; that is, the record is capable of delivering the message it was intended to communicate in order to achieve the purpose for which it was created." In other words, an authentic record is made up of content and metadata which is valid and correct, which as a whole, make the record complete and authentic. Let's take for example a medical record which contains attributes made up of content and metadata. The content consists of X-ray photos and metadata such as the name of the patient, the date and time that the X-ray was conducted, the equipment used to administer the X-ray, and a unique identification number to distinguish this record from other records similar to it. If any of the attributes of the example medical record went missing, became corrupt, or was incorrect, then the record as a whole would not be authentic and thus, not suitable as evidence to convey the health status of a patient.

Figures 1.1, 1.2, and 1.3, are examples of annual reports, governor's e-mails, and birth digital records. The contents of the digital records shown are displayed. As discussed later in this paper, during the transfer of these digital records, Kukini will add metadata, making the digital records complete. Note that the number of files

and the types of files for each of the digital records may vary.

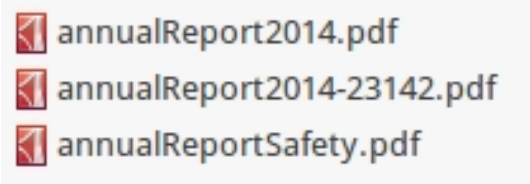


Figure 1.1: Annual Reports

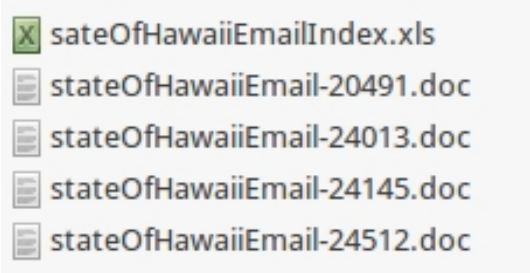


Figure 1.2: Governor's E-mails

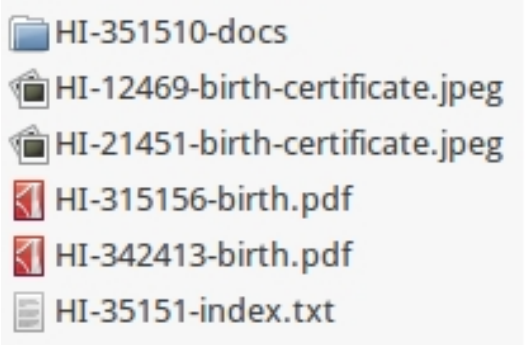


Figure 1.3: Birth Records

To summarize, a record can only be stated as authentic with high integrity if its content and necessary attributes are complete and correct. Although there are challenges in identifying a record as being authentic, there are also challenges in maintaining the authenticity and integrity of records. The following section will briefly discuss several of these problems, especially in the case when records are transferred between different states.

1.3 Risks that Affect the Authenticity of Records

Giaretta [16] states that "Digital records are at their point of greatest risk in "moments" when they are transitioning between states, e.g., when control is being passed to different systems. The authenticity of digital resources is threatened whenever they are exchanged between users, systems or applications, or any time technological obsolescence requires an updating or replacing of the hardware or software used to store, process, or communicate them."

An example of repercussions which occurred as a result of data transfer is in the case of The National Aeronautics and Space Administration (NASA)'s Apollo 11 Telemetry Data Recordings. According to NASA's The Apollo 11 Telemetry Data Recordings: A Final Report [27], the Apollo 11 moonwalk was filmed in slow-scan television (SSTV) format on telemetry tapes to be used as a backup in the event that the television broadcast recordings were to have failed. These telemetry tapes were transferred from Australia, to the Goddard Space Flight Center, to the Washington National Records Center, and recalled back to NASA, where the tapes were lost due to it being accidently purged and re-written to. NASA concluded that although the Goddard Space Flight Center and NASA followed all standard procedures for handling the telemetry tapes, this procedure was inadequate. They recognized the importance of having their current leaders ensure that documents and recordings of historical value make their way to the National Archives for preservation. If NASA and the Goddard Space Flight Center did successfully transfer the Apollo 11 telemetry tapes to the National Archives, would this have lead to the tapes being accessible and authentic in the 20th century? What procedures, structures, and processes are effective to ensure that records remain authentic and accessible over time? The following section will briefly discuss theories in regards to the processes and procedures

for maintaining the authenticity and integrity of records.

1.4 Maintaining the Authenticity and Integrity of Records

Organizations such the Hawaii State Archives seek to preserve, maintain, and make records accessible. These systems are often referred to as records systems. Duranti [7] stated that records systems, "preserve record structures, business context, and association with other like records. It preserves a record's authenticity (it is what it purports to be), reliability (accurate representation by a knowledgeable source), integrity (complete and unaltered) and usability (can be located, retrieved, presented, accessed, interpreted, and understood over time)." How do records systems achieve such a goal? Bekkers [9] stated that, "records can be imbedded with qualities and considered authentic when moved through time and space if the system and procedures are properly designed and this process of preservation begins at creation." What imbedded qualities must records consist of? What procedures and processes are needed? According to Thibodeau [14], digital records must possess the required attributes which include:

1. A stable content and a fixed form, meaning that the entity's content must be stored so that it remains complete and unaltered, and its message can be rendered with the same documentary form or presentation it had when first set aside.

2. Explicit linkages to other records within or outside the digital system, through a classification code or presentation it had when first set aside.

3. An identifiable legal-administrative, provenancial, and procedural context.

4. An identifiable author (i.e. the person or organization issuing the record), addressee (i.e., the person or organization for whom the record is intended), and writer (i.e. the person responsible for the articulation of content).

5. An action, in which the record participates or which the record supports either procedurally or as part of the decision making process.

6. A medium, that is a support or carrier which the record is affixed.

In order to preserve the authenticity of records, the procedures used within records systems must be properly designed so that the records processed are controlled with a high level of granularity; from records creation, through each and every state that the records are transferred through, until its final destination. As mentioned before, it is the goal of archives organizations to be able to preserve the authenticity of records and digital records. Duranti [7] states that most systems are mainly suitable for holding data, not preserving records. Do archives systems such as the Hawaii State Archives fit into this common category, even if their main mission is to preserve records? In the following section, this paper will discuss the Hawaii State Archives' need to upgrade their records system to increase their digital records preservation capabilities.

1.5 The Current Digital Records Preservation Capabilities of the Hawaii State Digital Archives

The Digital Preservation Capability Self-Assessment is a survey which measures an archive's current capabilities and services of preserving digital artifacts such as digital records. This survey is administered by the Council of State Archivists; a national organization who serve as directors of the principal archival agencies to ensure that the nation's documentary heritage is preserved and accessible. The Hawaii State Archives'

Digital Preservation Capability Self-Assessment, administered on July 2011, is partially shown in Figure 1.4, and is fully shown in Appendix E. This survey indicates that the current capabilities at which the Hawaii State Archives can preserve digital records are at "1 stage (Minimal)", meaning that, "Digital preservation capabilities are rudimentary and most electronic records that merit long-term retention are at risk."

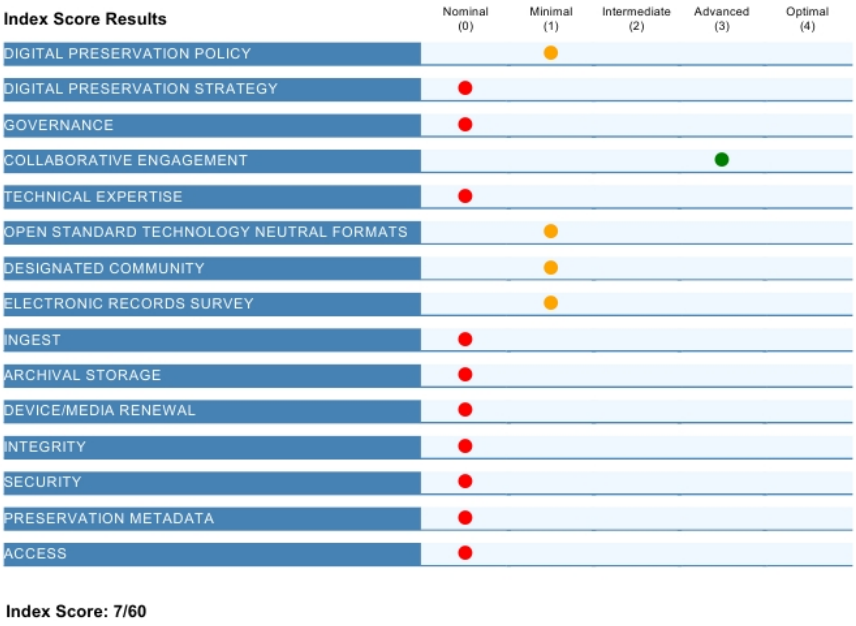


Figure 1.4: Digital Preservation Capability Self-Assessment

In order to improve their digital records preservation capabilities, the Hawaii Digital Archives is currently implementing a records system which consists of four major components. They include:

1. The development and proper use of a repository with the functionality to store and protect digital records of permanent value from loss, alteration, deterioration, and technological obsolescence;

2. The development of a search interface that will allow the public to enter specific search parameters, return records meeting those criteria and provide access to

those records, or parts of records, that are not restricted from public view by law.

3. The creation of the policies, procedures and documentation required to meet the criteria listed for certification as a trustworthy repository in accordance with the Trusted Repository: Audit and Checklist produced by OCLC-RLG (Online Computer Library Catalog Research Libraries Group, Inc.)

4. The development of a digital records transfer tool which must provide government agencies with the ability to transfer digital records to the Hawaii State Digital Archives. The transfer process must use secure and authenticated methods that document and ensure that the entirety of the files have been transferred uncorrupted.

The first point mentions the use of a repository to preserve digital records. This repository consists of a component called the, "the ingest pipeline". The ingest pipeline is made up of several components, which include databases to store the content and the attributes of records received, components to validate the integrity and authenticity of records received, notification components to notify, alert, and inform the records creator, records receiver, and other desired entities, and lastly, components to document the state of records as they pass through different phases and states of the ingest pipeline.

The search interface component will provide capabilities for digital records searching. The search parameters used to execute the search include the ability to search for records based on attributes such as the title of the record, the date that the record was created, the government entities from which it came from, the subject of the record, and more.

The certification granted by the OCLC-RLG, will allow the Hawaii State Archives

repository to be recognized by a qualified entity as being trustworthy; increasing their credibility and possibly usage.

The development of a digital records transfer tool which must allow government agencies with the ability to transfer digital records at a level of standard suitable for use by archival organizations such as the Hawaii State Archives. Because this component is the main topic of this paper, it is further discussed in its own section.

1.6 Kukini: A Digital Records Transfer Tool.

As mentioned in the previous sections, it is the goal of the Hawaii State Archives to update their records system to increase their digital records preservation capabilities. So the question remains, what qualities must a digital records transfer tool possess for it to be suitable for use within an archival framework? Re-stating the words cited by Giaretta [16], records are in their greatest moments of risk when they are being transferred between different states. Although there are risks involved, if a records system has been designed correctly to where the whole record, including its attributes are intact, and demonstrates a high level of control, then this suggests that it a system which is closely suitable within an archival framework.

Kukini is a Records Transfer Tool which has been implemented to meet the needs and requirements of the archival framework of the Hawaii State Digital Archives. Thus, it has been designed to:

1. Provide government agencies with the ability to transfer digital records to the Hawaii State Digital Archives.

2. Use secure and authenticated methods.

3. Document the transfer.

4. Ensure that the entirety of the files have been transferred uncorrupted.

The next section discusses the related work that has been done in the field, and the following section will describe the architecture and components that Kukini is comprised of and the features it provides.

CHAPTER 2
RELATED WORK

I analyzed several systems such as: version control systems, digital archives infrastructures, an archives related transfer tool, general purpose transfer tools, and power grid related transfer tools. I will summarize the general functionalities of each of the related systems mentioned and discuss their shortcomings. Although the authors of the papers discussed in this section provided effective designs, implementations, and solutions to their targeted environment, their tools do not meet the needs of the Hawaii State Archives. The Hawaii State Archives require a higher level of security, documentation, and control. Later in Chapter 3, I discuss Kukini, a solution which targets and satisfies several needs of the Hawaii State Archives in further detail.

2.1 Version Control Systems

Git and Subversion are version control systems used mainly for software configuration management. In software configuration management, digital files usually in the form of software code, are transferred, retrieved, or updated using commands issued to centralized and distributed repositories. The most commonly used Git commands are shown in Figure 2.1 Version control systems are analyzed as related work because Kukini contains a subset of the several functionalities that Git and Subversion provide.

Some of these functionalities include secure transfers over the network, and time-stamping of when a file was transferred or updated. Although these version control systems posses many desired functionalities required by the Hawaii State Archives, they are not suitable as transfer tool for use within an archival framework. For example, it has been developed to target software developers. It is mainly suitable

```
$ git help
usage: git [--version] [--exec-path[=<path>]] [--html-path] [--man-path] [--info
-path]
           [-p|--paginate|--no-pager] [--no-replace-objects] [--bare]
           [--git-dir=<path>] [--work-tree=<path>] [--namespace=<name>]
           [-c name=value] [--help]
           <command> [<args>]

The most commonly used git commands are:
   add        Add file contents to the index
   bisect     Find by binary search the change that introduced a bug
   branch     List, create, or delete branches
   checkout   Checkout a branch or paths to the working tree
   clone      Clone a repository into a new directory
   commit     Record changes to the repository
   diff       Show changes between commits, commit and working tree, etc
   fetch      Download objects and refs from another repository
   grep       Print lines matching a pattern
   init       Create an empty git repository or reinitialize an existing one
   log        Show commit logs
   merge      Join two or more development histories together
   mv         Move or rename a file, a directory, or a symlink
   pull       Fetch from and merge with another repository or a local branch
   push       Update remote refs along with associated objects
   rebase     Forward-port local commits to the updated upstream head
   reset      Reset current HEAD to the specified state
   rm         Remove files from the working tree and from the index
   show       Show various types of objects
   status     Show the working tree status
   tag        Create, list, delete or verify a tag object signed with GPG

See 'git help <command>' for more information on a specific command.
```

Figure 2.1: Commonly Used Git Commands

for transferring data and not records. It does not document the transfer process adequately with enough metadata.

2.2 Digital Archives Systems

Several digital archives systems presented by Nguyen, Askhoj, Smith, and Moore were analyzed [28] [8] [32] [26]. The authors presented several effective designs and implementations suited for their situation and their environment, but do not fulfill the specifications and requirements needed for the Hawaii State Digital Archives. For example, all of the digital archives infrastructures mentioned in these papers do not discuss their implementation, do not address the issue of validating and authenticating its users in order to ensure that they are trustworthy agents, and are platform specific.

The digital archives transfer tool called CINCH or the Capture Ingest Checksum,

developed by Rudersdorf [30] was analyzed. To use CINCH, users login and upload files through a web application. Several components such as duplication checking, virus scanning, validity checking, and metadata extraction are then executed. Once the execution process of these components have been complete, the files that the user uploaded are compressed into zip file(s). CINCH does provide several functionalities that Kukini possesses, but CINCH does not fully meet the requirements and specifications set forth by the Hawaii State Digital Archives. For example, CINCH only extracts metadata on the files that are uploaded and not metadata about the user's machine. Figure 2.2 displays the metadata that CINCH extracts.

Metadata Field	File Location Derived From
Author	The file header - embedded metadata
Creation date and time	The file header - embedded metadata
Last modified date and time	The file header - embedded metadata
Creator	The file header - embedded metadata
Producer	The file header - embedded metadata
File name	The original URL, stripped of special characters. May include a unique number as well, to avoid duplication.
Title	The file header - embedded metadata
Number of pages	The file header - embedded metadata (a null return indicates the number of pages could not be determined)
Subject	The file header - embedded metadata
Keywords	The file header - embedded metadata
Licensed to	The file header - embedded metadata
Possible title	Best guess, extracted from full text (if any)
Possible keywords	Best guess, extracted from full text (if any)
Checksum (SHA-1)	Computed by CINCH
Whether or not full text exists	Determined by CINCH

Figure 2.2: The Metadata that CINCH Extracts

CINCH is also only compatible to work with a limited number of file types which include: PDF, Microsoft Word, Excel, and PowerPoint, Jpeg, PNG, Gif, MP3, MP4, and Text files. This is limiting because some of the departments with the state of Hawaii have digital records which are made of TIFF files and other formats that are not supported by CINCH.

2.3 Power Grid Transfer Tools

Several general purpose and power grid related file transfer tool papers written by Meiss, Gu, He, Hanushevsky, and Sivakumar were analyzed [25] [17] [20] [18] [19] [31]. Several of these file transfer tools use UDP, which in nature, is an unreliable protocol, but do address this issue by re-sending the packets that were lost. During the transfer of digital records, ensuring that these digital records remain authentic and unaltered in any way is of major importance to the Hawaii State Digital Archives. Thus, we have decided that using UDP for file transfer will not meet our required specifications. These file transfer tools also focus on the issues of high speed performance and do not address the issues relating to authenticity and secure transfers.

2.4 File Transfer Protocol (FTP)

Several popular File Transfer Protocol (FTP) tools were analyzed. They include FileZilla, SmartFTP, and CoreFTP [1] [6] [2]. Although these tools do provide its users with the ability to transfer electronic files using secure and authenticated methods, they do not document their transfer process sufficiently. For example, Kukini will add additional metadata to the electronic files that it transfers. Kukini uses the bagit API [4] developed by the Library of Congress to package digital records in a format that is widely adopted by digital libraries. Kukini is also equipped with automatic update features which provides users with the ability to receive the latest features of Kukini over the network.

CHAPTER 3
DESIGN IMPLEMENTATION

I used several software tools and frameworks while working closely with State of Hawaii archivists, software developers, a senior software architect, a consultant, and several Hawaii State government agencies during Kukini's design, implementation, and evaluation. The following section discusses the software and tools used during Kukini's development. The next sections discuss the architecture and design of Kukini.

3.1 Development Environment

Kukini is implemented using the Java version 7 programming language. Kukini's code is managed with a Git source control repository, hosted on a Hawaii State Archives server. The Atlassian Software System is used for project management, issue tracking, collaboration, continuous integration, content sharing, and code quality. This includes build automation using Stash with the Maven build system. For each iteration of Kukini, its code and design were reviewed by the Hawaii State Digital Archive's software developers and senior software architect to ensure that the design and implementation of Kukini possesses a quality of standard which is satisfactory to the Hawaii State Archives. Much of the development environment was already configured and bought by the Hawaii State Archives, and thus, I was able to quickly access and use these tools as soon as I started developing Kukini.

3.2 Architecture and Design of Kukini

Kukini is an application which communicates with three different software components which includes an ApacheDS Lightweight Directory Access Protocol (LDAP) server, a Netbeans Update Center, and a Java Web Servlet called the SIP Transfer Servlet, shown in Figure 3.1. The role of each of these software components is discussed later in this paper. It is given now to provide a general idea of Kukini's architecture.

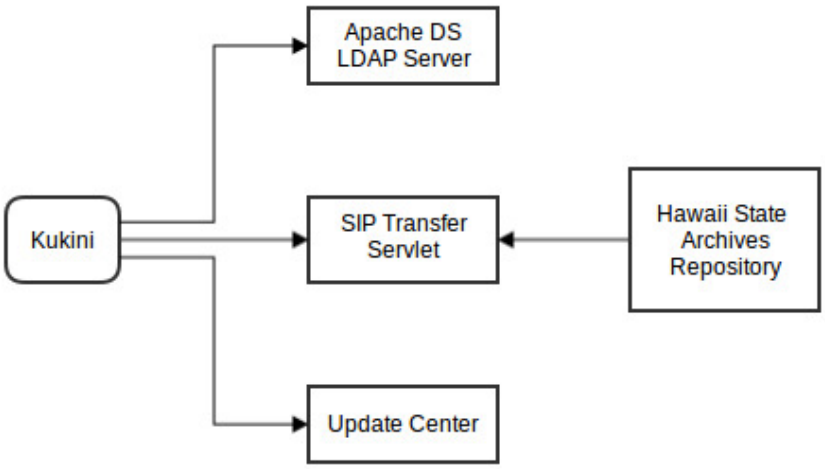


Figure 3.1: The Workflow of Kukini

Kukini is developed using the Netbeans Platform; a generic software framework. The Netbeans Platform Framework has been used extensively for the development of modular desktop applications. It provides several APIs such as a Window System for the creation menu items, toolbar items, keyboard shortcuts, Swing windows and other features. Modular applications are software systems that are made up of independent software components and are commonly referred to as "modules" by the software community. Kukini's architecture is composed of several modules which are shown in Figure 3.2.

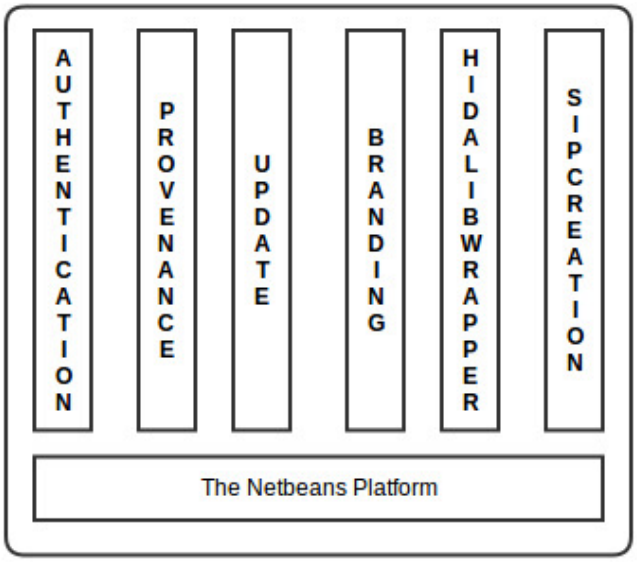


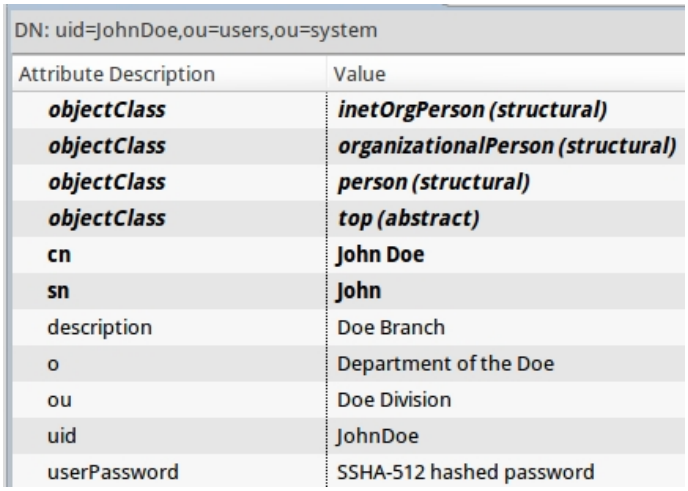
Figure 3.2: The Architecture of Kukini

Kukini's modules include the Authentication Module, the Provenance Module, the Update Module, the Branding Module, the HidaLibraryWrapper Module, and the SIPCreation module. Each module provides independent sets of features within Kukini such as an automatic updates feature, authentication mechanisms, and digital records transfer features. I used the Definitive Guide to the Netbeans Platform 7 textbook written by Bock as guidance during Kukini's development [10]. Each module is discussed in further detail in the proceeding sections of this paper.

3.2.1 The Authentication Module

The authentication module is a client which provides Kukini with methods to authenticate its users against Lightweight Directory Access Protocol(LDAP) servers. An ApacheDS LDAP directory server, which is referred to in this paper as the "Kukini LDAP Server (KLS)", is used for directory services such as storage, access, and authentication of users. The KLS contains user information such as, the username and

password used to login to Kukini, the user's given name and surname, and the government department, division, and branch of the user from which he or she is associated with. Figure 3.3 displays a sample user entry which was used to test the KLS. The Apache Studio software client was used to query and display the entry within this Figure. Due to security and confidentially reasons, actual user entries that are used in production are not described or shown in this paper.



DN: uid=JohnDoe,ou=users,ou=system

Attribute Description	Value
objectClass	***inetOrgPerson (structural)***
objectClass	***organizationalPerson (structural)***
objectClass	***person (structural)***
objectClass	***top (abstract)***
cn	**John Doe**
sn	**John**
description	Doe Branch
o	Department of the Doe
ou	Doe Division
uid	JohnDoe
userPassword	SSHA-512 hashed password

Figure 3.3: Kukini LDAP Server with Sample User Entry

The protocol that Kukini uses to communicate with the KLS is called the Transport Layer Security (TLS). TLS makes use of certificates and keys to establish a secure connection between client and server. Thus, the KLS is configured with a self-signed X.509 Secure Sockets Layer (SSL) server certificate and is TLS enabled. Using TLS, the KLS will only accept connections from clients containing a KLS shared certificate with its key. Likewise, Kukini will only connect to servers containing an authorized key and certificate which matches the KLS. If this handshaking process between Kukini and the KLS is successful, then a shared key is formed by both Kukini and the KLS, and is used to create an encrypted channel between them. Kukini then uses the Java Naming and Directory Interface (JNDI) to send queries to the KLS

in order to retrieve, update, or store user information using the encrypted channel. Users of Kukini login using the window shown in Figure 3.4. The Kukini login window is created using the Swing API.

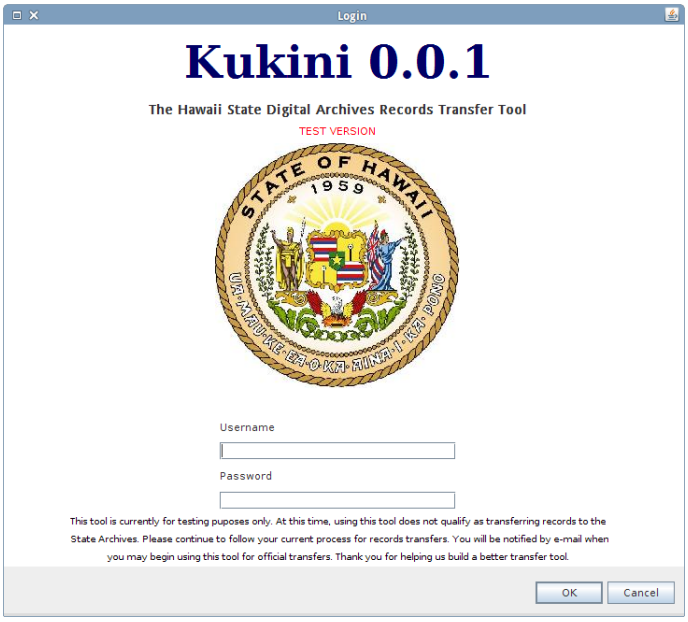


Figure 3.4: The Kukini Login Window

Once users have successfully logged in to Kukini, their department, division, branch, given name, and surnames are queried and are made available by other Kukini modules using the Netbeans Lookup API mechanism. Kukini uses this Lookup mechanism to access and use this user information as provenance metadata. The usage of this user information is discussed in the Provenance Module section of this paper.

In summary, the Authentication Module provides a mechanism to authenticate users of Kukini. This security mechanism decreases the chance of unauthorized users potentially transferring malicious files to the Hawaii State Archives repository. The Kukini Authentication Module increases the level of control in which digital records are handled and processed within the Hawaii State Archives records system.

3.2.2 The Provenance Module

The Provenance Module is used to extract provenance information about users transferring their digital records using Kukini. According to the Oxford English Dictionary [13], provenance is the chronology of the ownership, custody or location of a historical object, such as a digital record. Thus, if the provenance of a digital record is properly documented, this allows one to track the digital record's chain of custody and origins. The provenance information associated with a digital record can also assist in the verification of its authenticity. If for example, a digital record is sent from the Department of Agriculture, and its provenance information is read, showing that the record came from an unidentifiable IP Address, then this digital record would not be deemed as authentic by the Hawaii State Archives. The process of verifying the authenticity of digital records is not done by Kukini, but rather, by modules within the ingest pipeline. If the digital records sent are deemed inauthentic, then the Notification Module, which is also an ingest pipeline module, will log this event and send an e-mail to the use, describing the reason as to why the digital records sent was considered inauthentic. The inauthentic digital records are not kept for preservation, and are deleted. Users are then expected to fix the invalid digital records, and re-send them using Kukini.

The provenance information that Kukini extracts from a digital record is based on the Hawaii State Archives METS schema. According to the Library of Congress [5], "The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation." Some of the information contained within the Hawaii State Archives METS schema

include, the user information that is extracted by the Kukini authentication module, system information, and additional metadata such as the date of transfer, and information about the entity receiving the digital record (The Hawaii State Archives). This amalgamate of information is referred to as provenance information. Please see Appendix A for a full listing of the provenance information that is extracted.

To extract system information, Kukini uses an open source, cross platform API called the Hyperic's System Information Gatherer (SIGAR). SIGAR provides a Java interface to its native C libraries which allow the gathering of system information. The system information that is extracted includes the user's, ip address, mac address, operating system, operating system architecture, primary DNS, the username used to login to their machine, and more. This system information is then made available by other Kukini modules through the Netbeans Lookup API mechanism.

In conclusion, Kukini uses its Provenance Module in order to gather provenance information about its users. The provenance information of a record helps it to be identifiable as being authentic while providing additional attributes to a record; contributing to its completeness as a whole.

3.2.3 The Update Module

The Update Module provides users with a mechanism to automatically receive software updates, allowing them to download and install the latest features of Kukini remotely over a network. The Hawaii State Archives configured an Apache HTTP Server ("httpd"). This HTTP Server is used as the "Kukini Update Center (KUC)." The Update Module is a highly sought feature of the Hawaii State Archives, and thus was implemented during the timeline of this project.

The AutoUpdate Services API, provided by the Netbeans Platform, is the API used to develop the Update Module. Once users of Kukini have successfully logged in,

the Update Module is run in the background at Kukini's startup and queries the KUC remotely to retrieve the latest software features. The KUC is a server which hosts a list of modules in the form of NBM (Netbeans Module) files and an updates.xml file. This updates.xml contains a timestamp and lists all the NBMs stored within the KUC. Kukini reads this updates.xml file to determine whether there are any updates available. If the updates.xml contains a timestamp that is more recent than Kukini's last connection date, and possesses NBM version numbers that are greater than the modules of the user's current version of Kukini, then the NBM files from the KUC are automatically downloaded and installed. Figure 3.5, visually displays a state diagram of this update process.

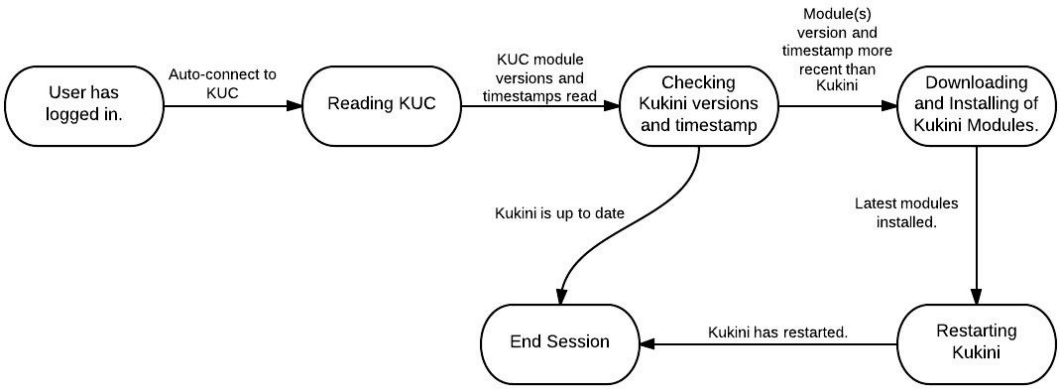


Figure 3.5: The Kukini Update Process

With the Update Module, users do not have to manually re-install newer executables of Kukini to use its latest software features, thus enhancing the user experience. Using the AutoUpdate Services API, users of Kukini can automatically disable or enable modules in relation to their identity. For example, if it is known that a user is from a certain department and they require a unique set of features, then these features are automatically installed and enabled just for this particular department. A TIFF reader, for example, may be a useful feature for the Department of Agriculture because some of their digital records consist of TIFF images, but may not be a useful

feature for the Department of Public Safety whose digital records mainly consist of PDF files. Also, if a user has logged in to Kukini and they have been detected to have an administrative role such as a records manager, then Kukini automatically enables modules to provide this user with user management features. For example, the ability to add additional user entries within the KLS. Although the enabling or disabling of modules based on identity is currently not in production, this feature has been tested by me, and thus could be easily distributed in production. Note that in the context of this paper, production means code that has been reviewed, and approved of by the Hawaii State Archives.

3.2.4 The Branding Module

The Branding Module provides the "look-and-feel" of Kukini. For example, fonts and font sizes, window names, window features, menu item names, icons, and the splash screen are configured and altered using the branding module. Kukini's splash screen, which is shown in Figure 3.6, is displayed at Kukini's startup. Note that the photo shown within the center of the splash screen, is credited to Gina Vergara of the Hawaii State Archives.

Figure 3.6: The Kukini Splash Screen

3.2.5 The HidaLibraryWrapper Module

The HidaLibraryWrapper (Hawaii Digital Archives Library Wrapper) Module is used to wrap and encapsulate jar files developed by the Hawaii Digital Archives and other 3rd party entities. By declaring a dependence on the HidaLibraryWrapper Module, Kukini's modules have access to their APIs. These APIs include the Spring Framework, a wrapper for the bagit API developed by the Library of Congress, database persistable model objects, an API to query tuple spaces, and the log4j API. The SIP Creation module uses the HidaLibraryWrapper Module to use the APIs that it provides.

3.2.6 The SIP Creation Module

The SIP Creation Module is one of Kukini's most important modules. It uses the Provenance Module and the HidaLibraryWrapper Module to securely transfer digital records through a network, documenting, and ensuring that the entirely of the files have been transferred uncorrupted. The expected question when confronted with the SIP Creation Module is, "what is a SIP?". This question is explained in the following paragraph.

The Submission Information Package (SIP)

According to Geller [15], under the Open Archival Information System (OASIS) Reference Model, a Submission Information Package (SIP) is material from a content provider that is transmitted to the archive. Content providers are the entities who create records. In the context of this paper, the Hawaii State Archives uses the term "SIP", to represent a single electronic file used to encapsulate the record series, of government agencies, with the goal of successful transmission to the Hawaii State

Archives repository. A record series is essentially a group of related digital records. For example, a medical record series would contain medical related records and birth record series would contain birth related records. There is a one-to-one relationship between a record series and a SIP. Thus a digital record(s), with its content, and provenance information are all contained with a single SIP. The structure of a SIP is disucssed in the next paragraph.

According to the OASIS model, SIPs are made up of one or many electronic records, where each electronic record consists of two parts, content: which are electronic files, and metadata: which is information related to those electronic files. For example, the content of digital records may consist of electronic files of type pdf, jpeg, text, html files, etc. The metadata part is the information which describes the content such as provenance relating to that content. Kukini uses the bagit API, developed by the Library of Congress, to create SIPs which conform to the bagit specification and in turn, also conforms to the OASIS SIP specification. The bagit specification revolves around the concept of the structure of a "bag." The structure of a bag consists of a directory called "data", a manifest file, a "bagit.txt" file, and optional "tag" and manifest tag files. The data directory contains the content of the digital record(s). The manifest file lists the content of the data directory as well as each content's checksums. This checksum can be used to check whether a SIP's content has become corrupted. The bagit.txt contains information describing the bag. The optional tag files are used to contain provenance about the contents of a bag. If tag files exist, then a manifest tag file is created within the bag. This manifest tag file lists each of the tag files and their checksums. An example of a SIP produced by the SIP Creation Module is shown in Figure 3.7. Placeholder names have been used to protect the record creator's identity. To create the data directory of the bag, Kukini re-creates the user's directory structure, where "C_Hida_Volume" represents the user's drive

volume from which the contents came from. For example, the file "<Placeholder file 1>.pdf" was stored on the user's machine at the path, "`C:/Users/<Placeholder dir>/Downloads/`". Further information about the files and directories of this SIP will become clear when the SIP creation process is explained in the next paragraph.

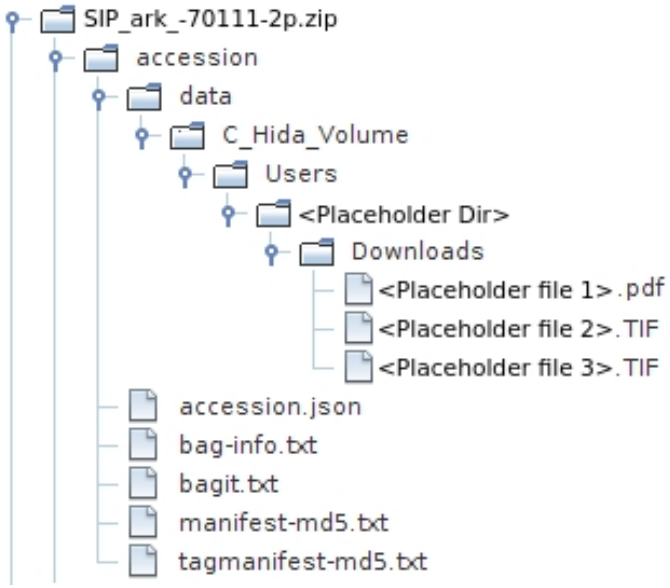


Figure 3.7: A Submission Information Package

SIP Creation Process

In production, before users are able to transfer their digital records using Kukini, a digital contract called the Records Transmittal Plan (RTP) is established between the users of Kukini and the Hawaii State Archives. For each type of record series being sent, an RTP will exist. There is a one-to-one relationship between the record series and its RTP. This RTP will contain information such as the file types, the files used to index the content, and the confidentially levels of the content within the record series. As of this writing, the current design of the RTP is still being discussed internally within the Hawaii State Archives. The following paragraphs will describe

the SIP Creation Module's current process in the creation of SIPs. In production however, users of Kukini will only be allowed to create SIPs based on their RTPs. Figure 3.8 visually displays the process of creating a SIP.

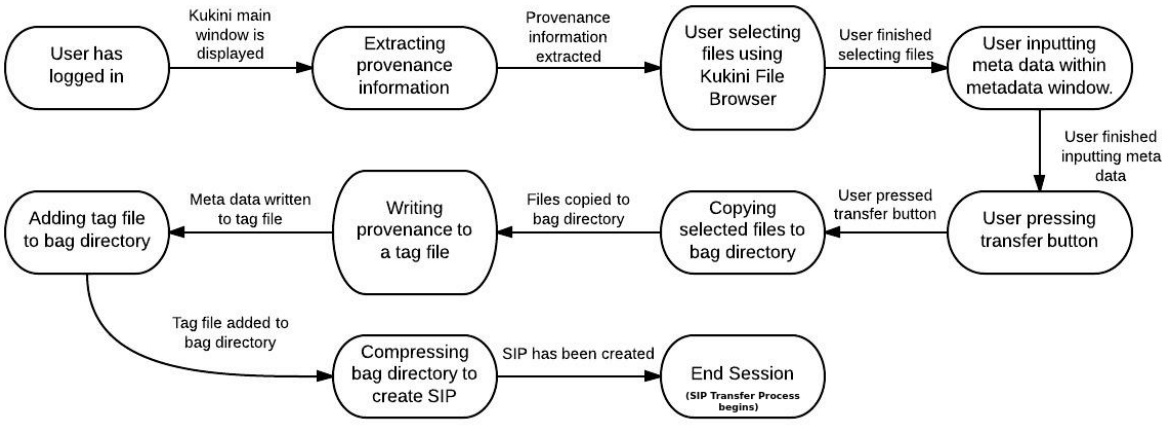


Figure 3.8: The SIP Creation Process

If the user logs in to Kukini successfully, the Provenance Module extracts the user's information and stores this information within memory. The user can then either select their electronic files representing the content of their SIP, or enter additional information using the Metadata window. The Metadata window is shown in Figure, 3.10. The Kukini File Browser, which is shown in Figure 3.9, is implemented from the Netbeans Platform favorites module and is used for file selection. Kukini's File Browser allows users to select files from their file system, external drives, and network drives.

As users select or deselect their files from the File Browser window, the full path of each file is listed or unlisted in the transfer window in real time. The transfer window is shown in Figure 3.11. This transfer window displays the electronic files that will be transferred once the transfer button is pressed. Please see Appendix D to view Kukini's Main Window which contains the File Browser, Metadata, and Transfer windows.

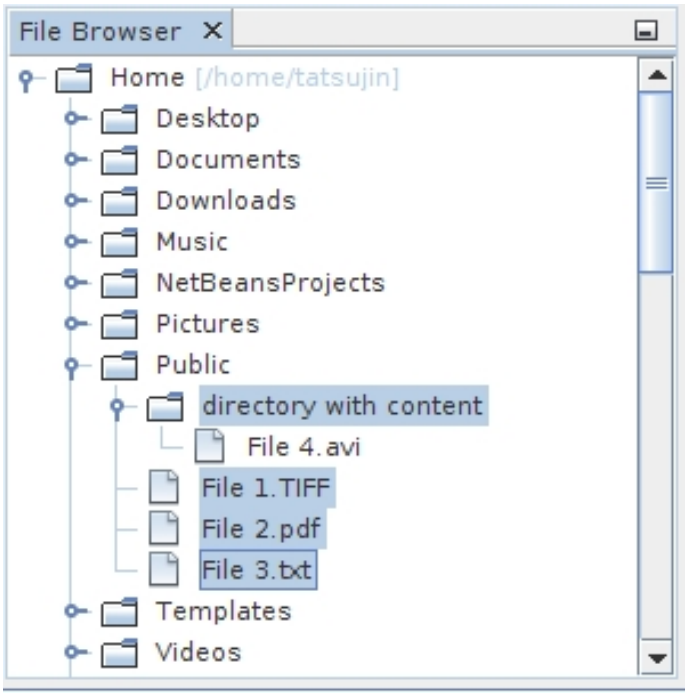


Figure 3.9: The Kukini File Browser Window

If the users of Kukini have finished selecting their files and have populated the metadata window with data, then pressing the transfer button will execute the final step in the creation of a SIP. First, the user and system information that was extracted from the provenance module is stored within a model object called an Accession. This model object, created by Dongie Agnir of the Hawaii State Archives, contains fields that are suitable for storing the Provenance Module's provenance information. For example, fields such as department, branch, and division are a part of the Accession model object. The Accession is then serialized in JSON format and written to a tag file called, "accession.json" within the SIP. Thus, the provenance information is written and can be read from the accession.json file. The content of the bag, represented by the files that the user has selected is placed within the bag, the manifests are created, and the entire bag is compressed in zip format. Using the Representational State Transfer Protocol (REST), the SIP is then transferred securely using HTTPS over

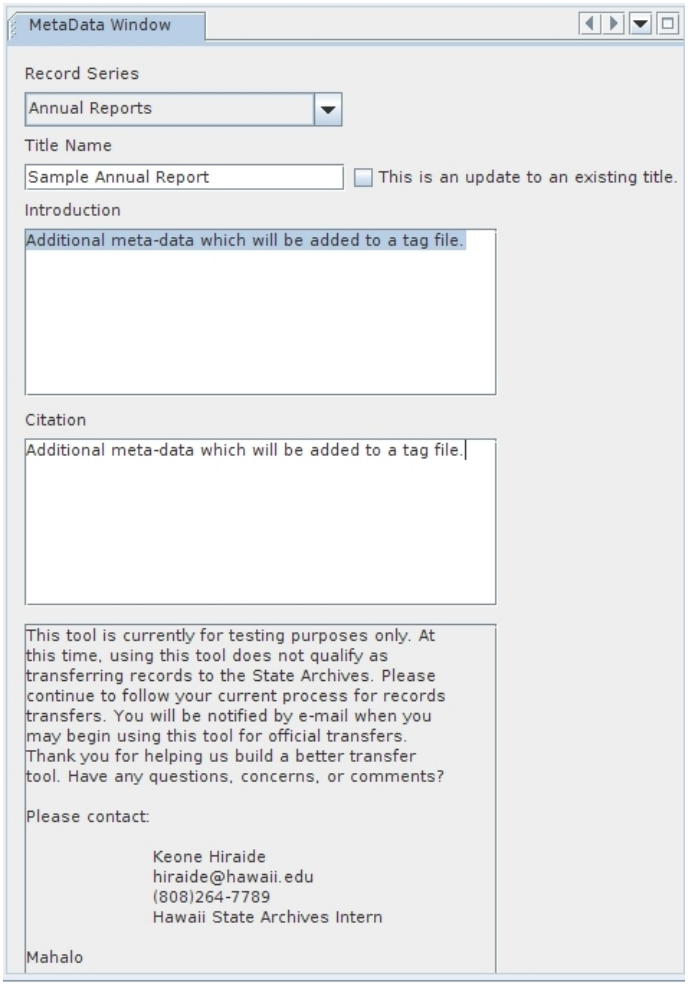


Figure 3.10: The Kukini Metadata Window

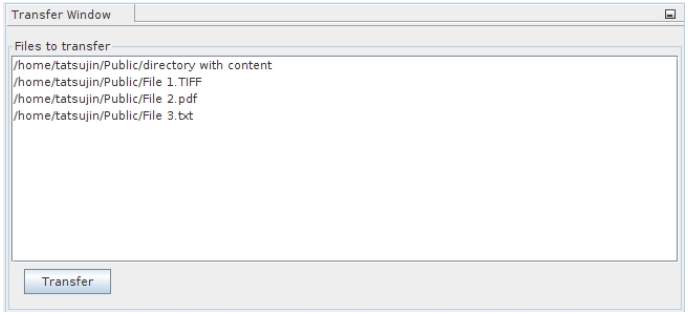


Figure 3.11: The Kukini Transfer Window

the network to the SIP Transfer Servlet, which was previously shown in Figure 3.1. This transfer process is discussed in detail in the next section of this paper.

SIP Transfer Process

In a client-server architecture, the SIP Creation Module is the client which communicates with the SIP Transfer Servlet (server). The SIP Transfer Servlet is a Spring Servlet deployed to an Apache Tomcat 7 server. The following paragraphs will explain the transfer and communication process between the SIP Transfer Servlet and Kukini's SIP Creation Module.

As mentioned in the previous section, the protocol used for the communication between the SIP Transfer Servlet and the SIP Creation Module is REST. Using REST, the state of resources such as Accessions are transferred between Kukini's SIP Creation Module and the SIP Transfer Servlet. Accessions are essentially SIPs which have been transferred from the SIP Creation Module and received by the SIP Transfer Servlet. In the event that the SIP is successfully transferred between the SIP Creation Module and the SIP Transfer Servlet, the SIP is referred to as an "Accession". This transfer process is called the "creation of an Accession".

Conforming to the REST protocol, Uniform Resource Locators (URLs) are used to locate resources. In the context of this paper, Accessions represent these resources. This URL not only locates Accessions, but also uniquely identifies Accessions, and serves equally well as a Uniform Resource Locator (URI) as it does a URL. Thus, the entire base URL can be used to uniquely identify Accessions within the Hawaii State Archives ingest pipeline. The URL to the SIP Transfer Servlet's accession resource, is shown in Figure 3.12. Using this base URL, clients can then create, locate, and retrieve the state of Accessions by sending REST requests to the SIP Transfer Servlet.

Kukini's SIP Creation Module uses Spring's RestTemplate API in order to send

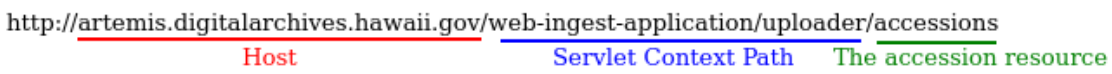


Figure 3.12: The SIP Transfer Servlet URL

HTTP GET and POST requests to the SIP Transfer Servlet. At the moment, the SIP Creation Module implements the RestTemplate's `getForEntity()` and `postForEntity()` methods to transfer the state of Accessions to and from the SIP Transfer Servlet. The `getForEntity()` method sends a HTTP GET request, returning an instance of type `ResponseEntity<Accession>` from the SIP Transfer Servlet. The `postForEntity()` method is used to create Accessions; and retrieves `ResponseEntity<Accession>` types from the server. These ReponseEntitys are further discussed in the following paragraphs.

The SIP Transfer Servlet uses Spring's resource-oriented MVC controller to process and handle REST requests. If the SIP Creation Module sends a request to the SIP Transfer Servlet, the response is an object of type `ResponseEntity<Accession>` . A `ResponseEntity<Accession>` instance encapsulates information about the request that was sent to the SIP Transfer Servlet. Information that can be retrieved from a `ResponseEntity<Accession>` instance which include: HTTP statuses, the returned body, and the location header. HTTP statuses indicate the status of the processed request. If an error occurred in the request's processing, then an error message can be retrieved from the ReponseEntity instance to learn about the cause of the error. The location header allows one to get the state of a created Accession, and is used for GET requests. As you may have noticed, the ResponseEntity being returned from the SIP Transfer Servlet has a generic type of "`Accession`". This means that a ResponseEntity's body will encapsulate an object of type `Accession`. By calling a `ResponseEntity<Accession>`'s `getBody()` method, an instance of type `Accession`

can be retrieved in JSON format, and deserialized into an `Accession` Object.

Table 3.2.6 visually encapsulates the SIP Transfer Servlet's POST and GET methods used to transfer the state of Accessions. The Use Case column lists the use cases that can occur if a POST or GET request for an Accession resource, to the request mapping "/accessions" is sent to the SIP Transfer Servlet.

Use Case	Request	HTTP Status
Successful creation of SIP	POST	201
The data uploaded was empty	POST	406
The data uploaded was invalid	POST	406
Failed to create the transfer directory	POST	500
Failed to write the SIP to the transfer directory	POST	500
Failed to mint an ID for the Accession	POST	500
Failed to write Accession to the transfer space	POST	500
Successful retrieval of an Accession	GET	200
Accession cannot be found	GET	404
Invalid request mapping.	POST/GET	404
Request mapping not supported	Not POST/GET	405

Table 3.1: Request Use Cases and HTTP Statuses

When users hit the Transfer Button using Kukini, a POST request is sent to the SIP Transfer Servlet. Figure 3.13 visually displays this process. In this figure, arrows without labels indicate an error that occurred in that state. This POST request transfers the SIP to the SIP Transfer Servlet over a network securely using HTTPS.

If a POST request on a SIP is successfully processed and handled, then the SIP Transfer Servlet places the SIP into a directory on its file system. As mentioned previously in this paper, this SIP would now be called an Accession because it has been transferred and retrieved by the SIP Transfer Servlet. This Accession is given a unique identification in the form of an Archival Resource Key (ARK) URI. Note that this ARK is what gets set in the location header within POST requests. Further provenance is then added to the Accession such as the date the Accession was received. This

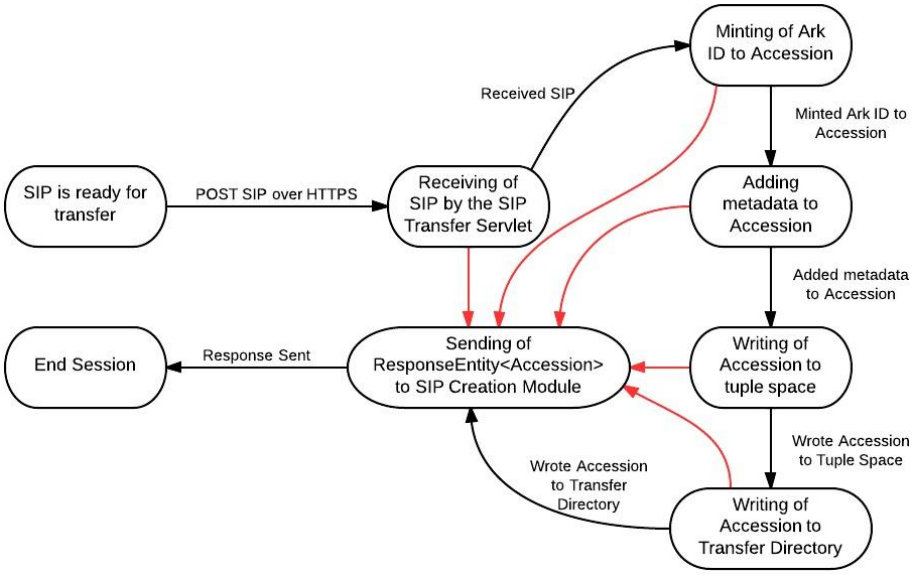


Figure 3.13: The Sending and Processing of a POST Request

Accession is then written to the Hawaii State Archives' tuple space for internal access. Lastly, the SIP Transfer Servlet will return a ResponseEntity<Accession>instance to the SIP Creation Module with its HTTP status set to 201(Created), its location header set to the ARK ID of the newly created accession, and a response body containing an instance of a `Accession` model object. If a POST request for a SIP is sent and a `ResponseEntity<Accession>` instance with its HTTP status set to 406(Not Acceptable) is returned, then this indicates that the request sent to the SIP Transfer Servlet contains an invalid resource. An invalid resource response occurs if the user sends a SIP with a file size of zero and/or is not of file type ".zip". If a POST request for a SIP is sent and a `ResponseEntity<Accession>` instance with its HTTP status set to 500(Internal Server Error) is returned, then this indicates that the server encountered an unexpected condition. These conditions can include: failure to mint an ARK ID for the Accession, failure to transfer the SIP resource to the SIP Transfer Servlet's transfer directory (`IOException` is thrown), and a failure to write the Accession to the tuple space. In all of the "failed" POST request cases, a HTTP response

is returned explaining the cause of the error within the message body as well as the specified HTTP status code set.

The SIP Transfer Servlet's GET requests are in the form of "/accessions/id" where the id parameter represents the Accession's ARK ID. Figure 3.14, visually displays the sending and the processing of a GET request.

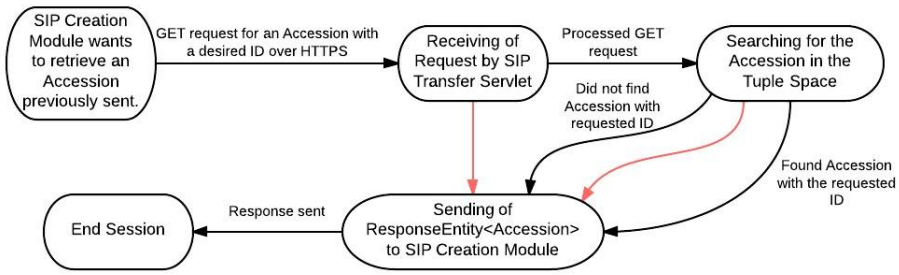


Figure 3.14: The Sending and Processing of a GET request

If a GET request for an Accession is successfully processed and handled, then an `ResponseEntity<Accession>` instance with its HTTP status set to 200(OK), its location header set to the ARK of the requested Accession, and its body mapped to an `Accession` instance. If a GET request for an Accession is sent and an instance of type `ResponseEntity<Accession>` with its HTTP status set to 404(Not Found) is returned, then this indicates that SIP Transfer Servlet could not find the resource that the client requested. Lastly, If a GET request for an Accession is sent and a `ResponseEntity<Accession>` instance with its HTTP status set to 500(Internal Server Error) is returned, then this indicates that the server encountered an unexpected condition. In all of the "failed" GET request use cases, a HTTP response is returned explaining the cause of the error within the message body as well as the specified HTTP status code.

If a POST or GET request with an invalid URL is sent to the SIP Transfer Servlet, then a `ResponseEntity<Accession>` instance with its HTTP status set to 404(Not

Found) is returned. Likewise, if a method other than POST or GET is sent to the SIP Transfer Servlet, then a `ResponseEntity<Accession>` instance with its HTTP status set to 404(Not Found) is returned.

In conclusion, the SIP Creation Module allows users to select their electronic records, package the contents and attributes of these electronic records into an electronic file called a SIP. This SIP is then securely transferred over the network to the SIP Transfer Servlet for further processing.

CHAPTER 4
USER EVALUATION

The User Evaluation chapter discusses the users and the processes that were involved in the testing and evaluation of Kukini.

4.1 The Focus Groups

The Hawaii State Archive's senior software developer and I conducted two focus groups on November 1st, 2013 to gather users for Kukini. 26 individuals from 13 departments participated in these focus groups. To increase the level of participation, we held two focus groups which consisted of 13 individuals each. The goal of this focus group was to meet the points of contact with the various government agencies that may eventually or are planning to use Kukini in production. The ultimate goal was to recruit their participation in testing, commenting, and providing feedback on Kukini as it evolved through development. We also asked them questions about their work environment as well as suggestions as to the features they would like in a Digital Records Transfer tool such as Kukini. Please see Appendix B, to read the notes that were gathered from both focus group sessions.

4.2 The Kukini Test Users

Out of the 26 individuals that attended the focus groups, 5 individuals were willing to be the testers of Kukini, and thus were recruited. The departments that these 5 testers came from included the Department of Agriculture, the Department of Accounting and General Services, the Department of Business, Economic Develop-

ment and Tourism, and the Department of Public Safety. Their titles included, Division Head, Planner, Planning Program Manager, Acting Planning Program Manager, Management Analyst, and Systems Analyst.

A consultant and 3 archivists from the Hawaii State Archives also participated in the testing and evaluation of Kukini. A meeting was conducted with this group to ask them questions about their work environment and if they had any immediate suggestions as to the features that they desired in Kukini.

Thus, a total of 9 test users participated in Kukini's testing.

4.3 Kukini User Testing

The first release of Kukini only contained the Branding Module and the Update Module. As you may recall from the Update Module section, this module provides Kukini with the feature of allowing users to receive the latest versions of Kukini remotely over a network. Kukini Windows and Linux installers were then created. I e-mailed, conducted phone calls, and spoke in person with 9 of my testers in order to schedule Kukini installation dates. I then met with each tester in person, and installed Kukini on their machines.

The immediate issue which was discovered at the time of installation was that the Update Module could not receive the latest Kukini features because of firewall issues. Thus, I contacted the ICSD department who maintains and has privileges to configure the government networks of the State of Hawaii. Once, the firewall issues were resolved, each of my test users could successfully receive the latest features of Kukini through the network as I made my latest software iterations and releases.

4.4 The Evaluation of Kukini

Once each module of Kukini reached a state of completion which is described in this paper, the System Useability Scale (SUS) questionnaire was administered to 3 testers who were archivists. According to Brooke [11], SUS is a reliable and effective way to evaluate the usability of systems such as Kukini. First, I described the various features of Kukini to them in an average timespan of 5 minutes and allowed them to ask me questions. I then gave them a single task to complete which was to: transfer digital files using Kukini. I left their cubicles, and let them fill the SUS questionnaire. The results show that archivist A scored Kukini at a 90, archivist B at 75, and archivist C at 70. The three Kukini testers also provided comments and suggestions about features they feel would improve the usability of Kukini. Although the sample size of this user evaluation was small, it revealed several weaknesses and strengths in regards to Kukini's current state.

4.5 The Kukini Evaluation Results

This evaluation revealed that users could successfully receive the latest features of Kukini using the KUS, could authenticate themselves with the KLS, and transfer electronic files over the network. I inspected their accession, and both the content and the attributes of their digital records were correct and valid. Thus, Kukini performed correctly according to its design. Analyzing the SUS scores of each of the three participates revealed that the category Kukini scored most poorly in was, "I needed to learn a lot of things before I could get going with this system." In regards to the user's comments, it was interesting to note that all three test participants suggested that a progress indicator of some kind be displayed to track the progress of digital records that are being transferred. Another universal suggestion amongst all

three participants were that they were not sure what to input in the "introduction" and "citation" fields within the metadata window. To see the full context of the participant's comments and suggestions, please see Appendix C. Lastly, the Hawaii State Archive's senior software developer also tested Kukini and pointed out areas in which Kukini can be improved. They are listed below:

1. Kukini does not restart when it says it will after updates. It just quits.
2. Branding is not updated to show the correct version. Should be 0.2.0 now.
3. Default NetBeans options/ keymap hasn't been pruned. Shouldn't have all of that irrelevant junk in there.
4. Options/misc/files pane should be tailored for Kukini.
5. User needs to be able to navigate outside of their home directory.
6. Tab names should not have the word "window" in them.
7. In the window menu, the File Browser window is mislabeled "Favorites". (What is the badging on that menu item?)

The comments and suggestions communicated by the testers and the senior software developer have been added to the Hawaii State Archives Jira backlog. Thus, software iterations to satisfy these comments and suggestions will be resolved within the next two week sprint of this writing. Because I am hired by the Hawaii State Archives, I will be the developer making the concrete changes to Kukini to address its problems. These changes are discussed in the future work section of this paper. Overall, users were satisfied with the current state of Kukini as a digital records transfer tool.

CHAPTER 5
CONCLUSION

In conclusion, I have developed a digital records transfer tool which provides government agencies with the ability to transfer digital records to the Hawaii State Digital Archives. The transfer process uses secure and authenticated methods that documents and ensures that the entirety of the files have been transferred uncorrupted. During the development process of Kukini, software developers critiqued and reviewed its code and design in code reviews.

As mentioned section 1.4 of this paper, in order to validate records as being authentic, both the content and the attributes of a record must be intact and correct. A properly designed Records systems can preserve the authenticity digital records, even if they have been transferred through time and space. Some of the traits which suggest a properly designed record system include the ability to handle/process digital records with a high level of control starting at records creation. Kukini provides this high level of control with the use of its modules, and is suited for handling both the content and attributes of digital records.

To test and evaluate Kukini, I gathered government agencies and archivists to recruit them as software testers. Once the Update Module was completed, I installed Kukini on each of the user's machines and worked on Kukini until each of its modules were completed to its current state.

Lastly, based on Kukini's user evaluation using the SUS questionnaire, users of Kukini made suggestions and comments in regards to Kukini's current state. These comments and suggestions were added to the Hawaii State Archive's backlog for reference and completion. The concrete changes that will be made to Kukini to address its problems is discussed in the future work section of this paper. Because

I have been hired by the Hawaii State Archives, the changes discussed in the future work section will mostly likely be implemented by me in the near future.

5.1 Future Work

This section discusses in my opinion, in the order of their importance, the features that can be implemented to Kukini to improve its quality and usability. I begin by discussing the most important features first and end in the feature I feel is least important.

5.1.1 The Software Architect's Suggestions

Kukini will be updated to fix the problems that the senior software architect mentioned. He stated that Kukini failed to restart once an update occurred. 11 machines successfully completed Kukini's update process, ending with Kukini correctly restarting. Since this seems to be an outlier case, the machine specifications that the senior software architect used to update Kukini has been noted. Since am unable to replicate this problem, I may have to use the software architect's machine to do so. I will continue to analyze this problem until it is resolved.

To resolve Kukini's branding problems, updates will be made to display the correct version number within the splash screen and the login window. The word "Window" will be removed within Kukini's File Browser, Metadata, and Transfer window tabs. Within the window menu, the File Browser will be correctly re-labled to "File Browser", and the badging will be explained.

In regards to the options/keymap feature, a user expressed that the keymap feature was useful, and thus whether it should be removed or not will be further discussed internally within the Hawaii State Archives.

In regards to the user's ability to navigate outside of their home directory, this feature has already been implemented as of this writing. This suggestion further reenforces the need for a tutorial, wizard, user guide, or help icons to help explain the capabilties and features of Kukini.

5.1.2 Additional Provenance Information

Additional provenance information could be added to the electronic records being transferred. This provenance information includes the version of Kukini that the user used to transfer files, the full path of each file, the user permissions of each file, and the last modification dates of each file. This provenance information will help in verifying the authenticity of the digital records that are sent using Kukini. For example, if a user were to send digital records using a version of Kukini that is not valid due to it being outdated or incorrect, then this information would be known through the reading of the digital record's provenance information. Embedding Kukini with a hashed version number could also prevent users from sending digital records using an unauthorized version of Kukini; the SIP Transfer Servlet will only accept digital records from authorized versions of Kukini.

5.1.3 Dashboard

After talking with several archivists and software developers at the Hawaii State Archives, the creation of a "dashboard" within Kukini with the functionality to display the status of user Accessions as they are being processed within the Hawaii State Archives ingest pipeline is very valuable. Figure E.1 is a prototype of the Kukini dashboard. This dashboard will also serve as a "history window", which was suggested by the users of Kukini. Using this dashboard, users will know whether the current

state of their Accession is being scanned for viruses, being checked for corruption, or deemed valid/invalid. This dashboard would serve as reliable record in itself; providing users with a receipt of the digital records that they have previously sent. Since Kukini already supports the sending of GET requests, accessions can be retrieved, and its attributes can be displayed within a separate window. This separate window will eventually become the Kukini dashboard.

5.1.4 Transfer Progress Notification

A progress bar to visually show the progress of a transfer will be implemented. The Netbeans Platform provides an API to visually display the progress of tasks. Thus, a progress bar can be created to display the user's transfer process. The tasks of this transfer will include, the packaging of the user's electronic files into a SIP, the extraction of the provenance information, the serializing of this provenance information, the writing of the tag file within the SIP, then finally, the transfer of this SIP to the SIP Transfer Servlet.

When the user's electronic files are being transferred, badging will be added to the transfer window to display its progress. For example, a "green check" icon to indicate the successful transfer of files and a "red X" to indicate the failure of transfer.

5.1.5 SIP Creation Assistance

An original goal of Kukini was to have it "assist" users during their digital file selection process. For example, Kukini's File Browser could highlight files that are named a certain way. This feature may be useful when it is known that a department uses a strict naming convention to name their digital files. The File Browser could also limit the files that are shown based on their file types or even providing "badging" to

indicate important files.

5.1.6 Tutorial

To train the users of Kukini about its features, a wizard will be created using the Netbeans Platform's Wizard API. This wizard will guide the user through a digital records transfer process. For example, the first window could teach users about the selection of files using the File Browser. The second window could explain the metadata window and its expected input, etc. When users hover their cursor over certain GUI components within Kukini, a bubble with text will appear. This text will describe the GUI component's functionalities. A user manual will also be created to help supplement the learning process. This user manual will not contain implementation details, but rather, explain Kukini's features in the form of a tutorial, targeted towards its users.

5.1.7 Ingest Features Within Kukini

Another goal would be to eventually integrate features within Kukini which are currently a part of the ingest pipeline. Some of these features may include virus scanning, duplication checking, FITS characterization, the validation checking of digital records and notification.

5.2 Contributions

I claim the following contributions:

1. An open source extensible digital records transfer tool system.

2. A system that is well documented.

3. A system which fits within an archival framework and is freely available for use by other archival organizations.

4. A project which has established a user base consisting of several Hawaii government entities.

I also claim that with Kukini, the digital preservation capabilities of the Hawaii State Archives has increased. If the Council of State Archivists were to assess the Hawaii State Archives digital preservation capabilities with Kukini, I feel that the score of the Hawaii State Archives would increase in the areas of governance, technical expertise, ingest, integrity, security, and preservation metadata. Figure 5.1 visually displays my opinion and my prediction of how the Hawaii State Archives Digital Preservation Capability Self-Assessment will look as a result of Kukini's contributions.

Figure 5.1: Digital Preservation Capability Self-Assessment with Kukini

I feel that governance will increase because the identity of the various stakeholders and their roles are known when users login using Kukini. The department, division,

branch, full name, and other related information of the stakeholders are extracted by Kukini's Authentication Module.

The technical expertise has been increased within the Hawaii State Archives. Kukini supports project-based digital preservation initiatives and supports conforming Submission Information Packages and Archival Information Packages.

The ingest digital preservation capabilities has been increased as a result of Kukini. Fully conforming SIPs are transferred and ingested automatically. Once received, they are automatically checked for viruses, format validations, and finally stored.

The integrity of digital records received has been increased because of Kukini. Using the bagit API, Kukini generates an electronic package which produces an MD-5 hash of each electronic file being transferred. The MD-5 digests of digital records are intact before and after they are received by the Hawaii State Archives repository.

The level of security in which digital records are handled has been increased as a result of Kukini. Digital records are transferred through an encrypted channel to the Hawaii State Archives repository. The Hawaii State Archives repository is protected through comprehensive firewall protection and is equipped with comprehensive role based access rights management.

The preservation metadata extraction capabilities of the Hawaii State Archives has improved. Kukini's metadata extraction is based off of a PREMIS-based preservation metadata schema which includes a data dictionary that identifies ISO 14721 conforming comprehensive preservation metadata specifications.

Thus with Kukini, the digital preservation capabilities of the Hawaii State Archives is predicted to increase by 13 points, totaling 20/60 points. In order to reach the maximum score of 60 points, further work is needed in respect to all of the Hawaii State Archive's records system components. These components include the digital records repository, the search interface, the creation of policies and procedures, and Kukini.

APPENDIX A
PROVENANCE INFORMATION EXTRACTED BY THE PROVENANCE MODULE

1. User Information
    (a) Common name
    (b) Department
    (c) Divison
    (d) Branch
2. Network Information
    (a) Network interface name
    (b) Ip address
    (c) Mac address
    (d) Net mask
    (e) Host name
    (f) Domain name
    (g) Default gateway
    (h) Primary DNS
    (i) Secondary DNS
3. System Information
    (a) System description
    (b) System patch level

(c) System vendor

(d) System vendor code name

(e) System vendor version

(f) System version

4. METS Information

(a) Name of the Preserver

(b) Date of transfer

(c) Date of Creation

(d) Security and control procedures used for the transfer

(e) System vendor version

(f) System version

APPENDIX B
NOTES GATHERED FROM FOCUS GROUP SESSIONS

These are notes gathered from both focus group sessions of Kukini. These are points that some of the attendees stated at the meetings. The department and the names of the speakers will not be revealed for confidentially reasons.

1. Many records are still in paper form.
2. Possesses records which are completely open to the public and shown on their website.
3. Need to scan paper records to make them digital.
4. Would like to enter their own metadata suitable for their department.
5. Too time constrained to be a tester of Kukini at the moment.
6. Partners with the FBI to ensure that records are locked and secured. These records are very confidential.
7. Not clear about the differences between data and records.
8. Some records should never be deleted because they are evidence.
9. Started producing digital records since 2008.
10. Not clear as to what permanent records are.
11. Have records containing boat names, which are confidential.

APPENDIX C
SYSTEM USABILITY SURVEY

The System Usability Survey Consists of 10 questions. For those 10 questions, users can give a score of either: Strongly Disagree, Disagree, Neutral, Agree, and Strongly Agree.

C.1 Kukini Evaluation by Person A

Person A was an archivist employed by the Hawaii State Archives.

C.1.1 Person A SUS Evaluation

1. I think that I would like to use this system frequently.

   Neutral

2. I found the system unnecessarily complex.

   Disagree

3. I thought the system was easy to use.

   Strongly Agree

4. I think that I would need the support of a technical person to be able to use this system.

   Disagree

5. I found the various functions in this system were well integrated.

   Agree

6. I thought there was too much inconsistency in this system.

   Disagree

7. I would imagine that most people would learn to use this system very quickly.

   Agree

8. I found the system very cumbersome to use.

   Strongly Disagree

9. I felt very confident using the system.

   Agree

10. I needed to learn a lot of things before I could get going with this system.

    Disagree

Person A Suggestions and Comments

Person A had these comments and suggestions in regards to Kukini.

1. For first time users - need explanation or examples on what is expected on the blanks:

- Introduction
- Citation
- Title Name

2. (Would be) nice to have visual feedback that files are being transferred. A timer or any other visual... Otherwise I don't know if it is working and I will click on the transfer button again.

C.2 Kukini Evaluation by Person B

Person B was an archivist employed by the Hawaii State Archives.

C.2.1 Person B SUS Evaluation

1. I think that I would like to use this system frequently.

   Strongly Agree

2. I found the system unnecessarily complex.

   Disagree

3. I thought the system was easy to use.

   Strongly Agree

4. I think that I would need the support of a technical person to be able to use this system.

   Strongly Disagree

5. I found the various functions in this system were well integrated.

   Strongly Agree

6. I thought there was too much inconsistency in this system.

   Strongly Disagree

7. I would imagine that most people would learn to use this system very quickly.

   Strongly Agree

8. I found the system very cumbersome to use.

   Strongly Disagree

9. I felt very confident using the system.

   Strongly Agree

10. I needed to learn a lot of things before I could get going with this system.

    Agree

Person B Suggestions and Comments

Person B had these comments and suggestions in regards to Kukini.

1. Tools >Plugins >Installed (tab)

   Add a plugin history with possibly dates for troubleshooting purposes.

2. Metadata Window

   1. Include a clickable help icon next to each fieldname so users have an idea what goes into the field. Would help facilitate ease of use without tutorial. i.e. I don't know what the citation field is for.

   2. Once I expand it to a certain size, I can no longer resize it smaller. If the checkbox is causing this, maybe that can go under the title field.

3. Transfer History

   1. Perhaps a seperate window that tracks what has been transferred. The transfer window may list items that are currently transferring but if I wasn't sure if I had already transferred a file, the history could be helpful. Allow the history to list per transfer job, when the user clicks on it, the metadata they associated with the transferred files, as well as the list of what was transferred pops up. They can refer to the history for the same metadata used when transferreing new files within the same record series.

4. Transfer Window

    1. Any way to visually allow the users to see that it is in the process of transferring. Upon clicking, I don't get feedback and am liable to keep clicking on the "Transfer" button repeatedlly. Maybe a check button next to each file as it is transferred? A progression bar?

5. Quick Wizard

    1. Wizard to do basic transfer might be nice. Sometimes people don't like to read. Allow people to run it again if they go to the help window.

C.3 Kukini Evaluation by Person C

Person C was an archivist employed by the Hawaii State Archives.

Person C SUS Evaluation

1. I think that I would like to use this system frequently.

    Strongly Agree

2. I found the system unnecessarily complex.

    Disagree

3. I thought the system was easy to use.

    Agree

4. I think that I would need the support of a technical person to be able to use this system.

    Agree

5. I found the various functions in this system were well integrated.

   Agree

6. I thought there was too much inconsistency in this system.

   Disagree

7. I would imagine that most people would learn to use this system very quickly.

   Agree

8. I found the system very cumbersome to use.

   Disagree

9. I felt very confident using the system.

   Agree

10. I needed to learn a lot of things before I could get going with this system.

    Agree

Person C Suggestions and Comments

Person C had these comments and suggestions in regards to Kukini.

1. In case of operator error - accidently moving folder to transfer box

   How can accidently placed file be removed? I exited the program to prevent accidental transfer.

2. Not sure what to include in title name, introduction, and citation.

3. Metadata Window
   1. Explanation box - What is metadata?
   2. Dialog, help - Why is metadata important to sender, receiver, and researcher.

4. Process Bar so we don't keep clicking transfer.

5. Notification E-mail

APPENDIX D
THE MAIN WINDOW OF KUKINI

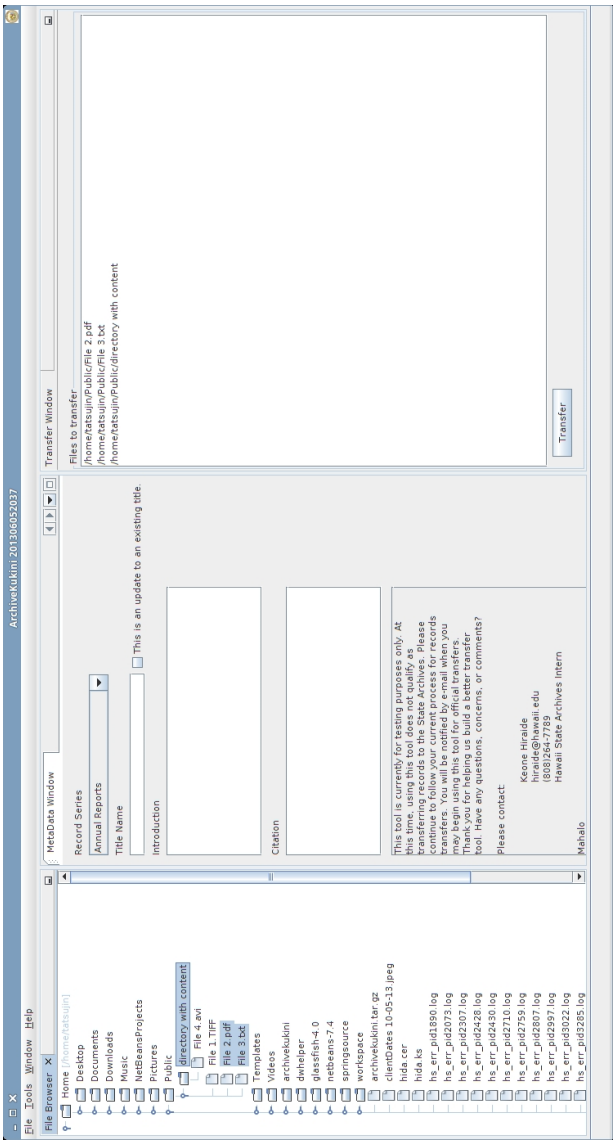


Figure D.1: The Main Window of Kukini.

APPENDIX E
DASHBOARD

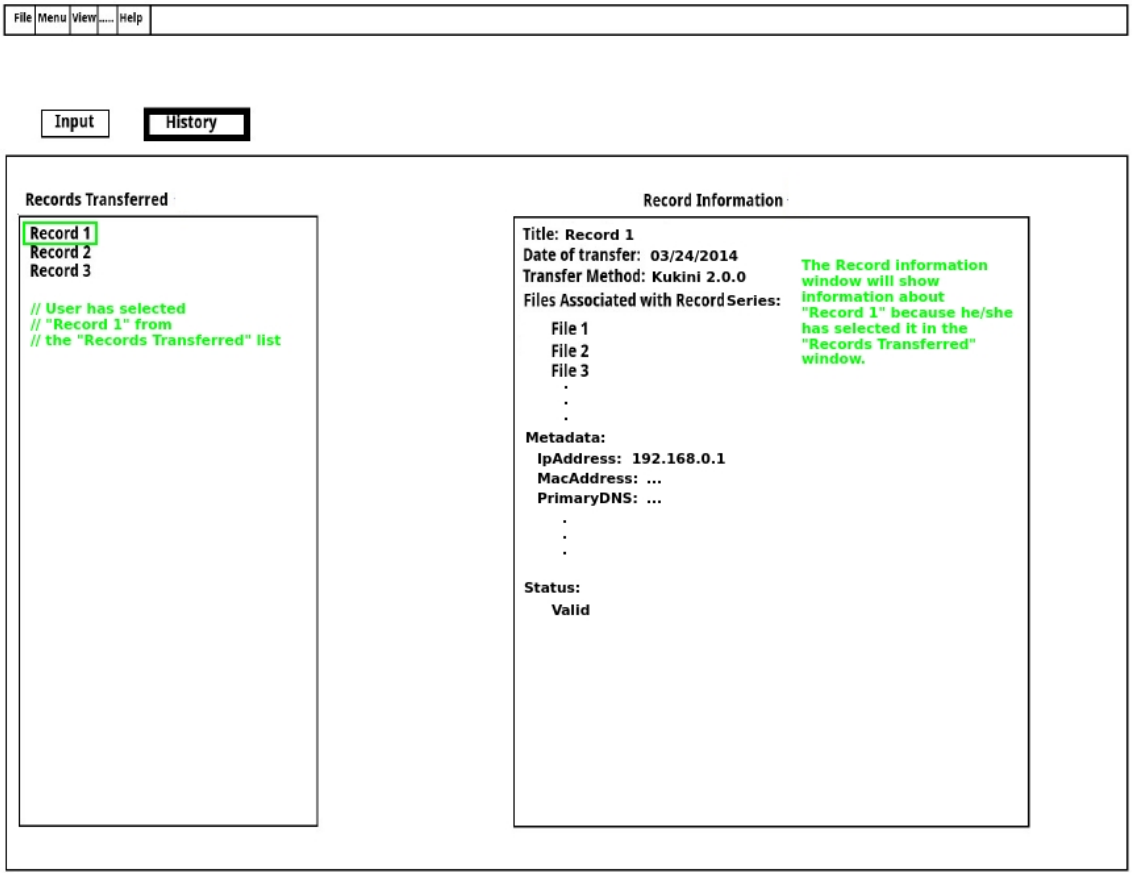


Figure E.1: Dashboard Prototype of Kukini.

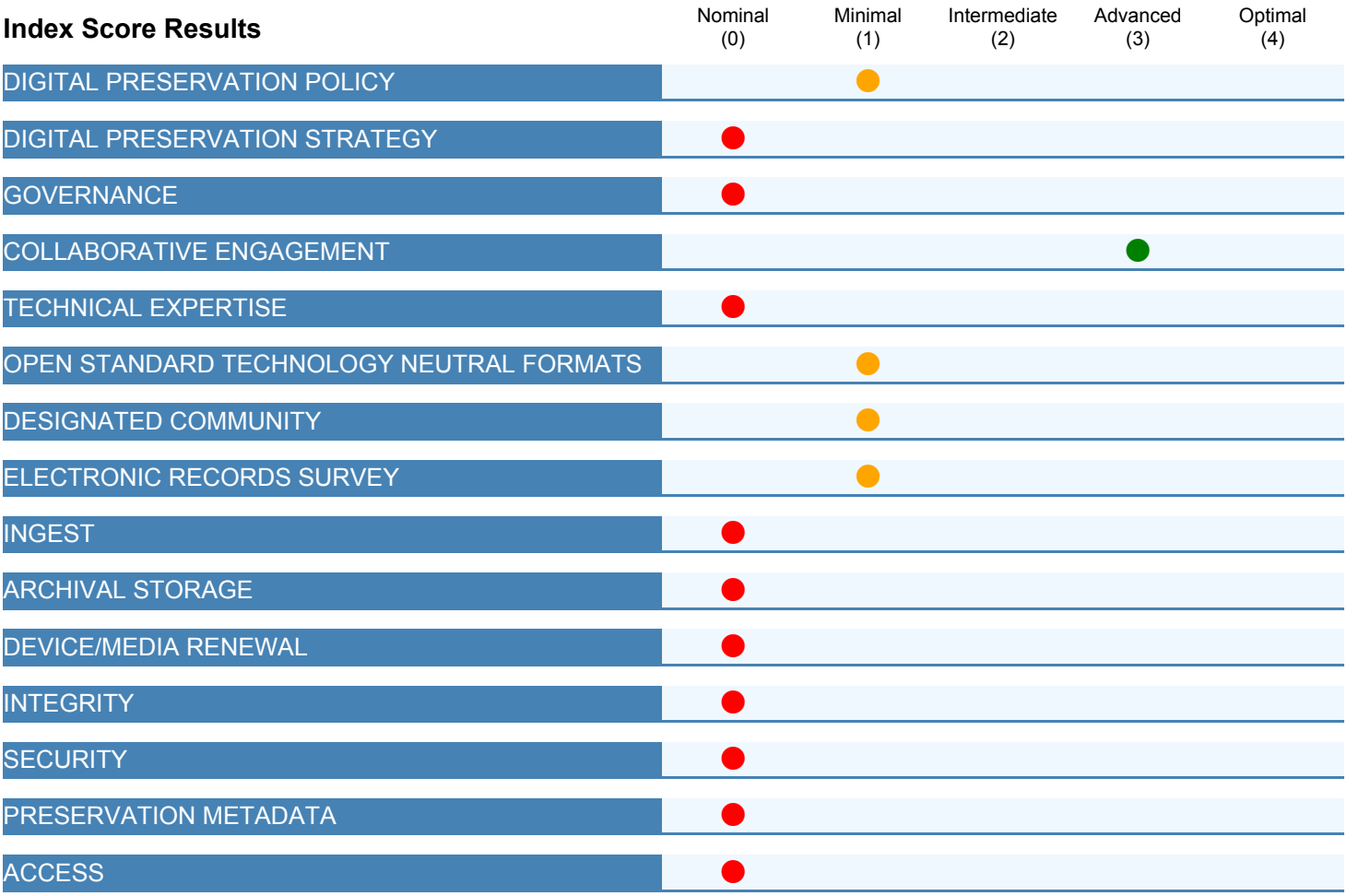
APPENDIX F
DIGITAL PRESERVATION CAPABILITY SELF-ASSESSMENT

The Digital Preservation Capability Self-Assessment is a survey which measures an archive's current capabilities and services of preserving digital artifacts such as digital records. This survey is administered by the Council of State Archivists; a national organization who serve as directors of the principal archival agencies to ensure that the nation's documentary heritage is preserved and accessible. The Hawaii State Archives' Digital Preservation Capability Self-Assessment, administered in July 2011, is fully shown in this appendix and indicates that the current levels at which the Hawaii State Archives can preserve digital records are at "1 stage (Minimal)", which means that, "Digital preservation capabilities are rudimentary and most electronic records that merit long-term retention are at risk.

Digital Preservation Capability Self-Assessment

Name:	Hawaii SERI Self-Assessment Team
Title:	Hawaii State Digital Archives
Agency:	Hawaii State Archives
State:	HI
Reports To:	Department of Accounting and General Services
Contributors:	Susan Shaner, State Archivist Gina Vergara-Bautista, Digital Archives Project Manager

Index Score Results



Index Score: 7/60

Based upon your responses, the digital preservation capabilities and services of your archive/records management unit falls into the **1 Stage (Minimal)**. Digital preservation capabilities are rudimentary and most electronic records that merit long-term retention are at risk.

This scorecard indicates the current capabilities of the Archives/RM unit for each component in the Digital Preservation Capability Maturity Model. The filled in circles (red, orange, yellow, light green, dark green) denote where all of the respective requirements have been met.

1. DIGITAL PRESERVATION POLICY

The government unit charged with ensuring preservation and access to permanent legal, fiscal, operational, and historical electronic records should issue its digital preservation policy in writing including the purpose, scope, accountability, and approach to the operational management and sustainability of trustworthy digital repositories.

- ☐ The Archives/RM unit does not have a written digital preservation policy.
- ☑ The Archives/RM unit has a digital preservation policy in development but it has not yet been approved or issued.
- ☐ The Archives/RM unit has issued a digital preservation policy and it is widely disseminated to stakeholders.
- ☐ The Archives/RM unit periodically conducts a self-assessment and reports its adherence to its digital preservation policy.
- ☐ The Archives/RM unit reviews and updates the digital preservation policy at least every two years.

2. DIGITAL PRESERVATION STRATEGY

The organization charged with the preservation of permanent electronic government records must proactively address the risks associated with technology obsolescence including plans related to periodic renewal of storage devices, storage media, and adoption of preferred preservation file formats.

- ☐ The Archives/RM unit does not have a plan to address technology obsolescence.
- ☐ The Archives/RM unit accepts electronic records in their native format on an ad hoc basis with the expectation that new software will become available to support these formats.
- ☑ The Archives/RM unit encourages records producers to retain records of long-term value in preservation-ready file formats.
- ☐ The Archives/RM unit proactively and systematically monitors changes in technologies that may impact the digital records collections and the archival repository.
- ☐ The Archives/RM unit implements the transformation of selected native file formats to preferred/supported preservation file formats in the archival repository.
- ☐ The Archives/RM unit implements transformation of all electronic records from records producing units to preferred preservation file formats in the archival repository.

3. GOVERNANCE

The state or territory has a formal decision-making framework that assigns accountability and authority for the preservation of electronic records with permanent historical, fiscal, operational or legal value, and articulates approaches and practices for trustworthy digital repositories sufficient to meet stakeholder needs. Governance is exercised in conjunction with information management and technology functions and with other custodians and digital preservation stakeholders such as records producing units and records consumers and enables compliance with applicable laws, regulations, record retention schedules, and disposition authorities.

- ☑ The state/territory does not specifically address digital preservation requirements in the scope of current governance activities.
- ☐ A project- based digital preservation governance framework is operational or has been successfully completed.
- ☐ The operational state/territory-wide digital preservation governance framework identifies the various roles of stakeholders in the preservation of electronic records.
- ☐ An operational state/territory-wide digital preservation governance framework is in place that assigns accountability and authority for the preservation of electronic records.
- ☐ The state/territory digital preservation governance framework specifies an on-going commitment to the sustainability of an ISO 14721 conforming archival repository.
- ☐ The operational state/territory-wide digital preservation governance framework for digital preservation is reviewed and updated at least every two years to take into account changing technologies and new organizational structures.

4. COLLABORATIVE ENGAGEMENT

Digital preservation is a shared responsibility so the organization with a mandate to preserve electronic government records in accordance with accepted digital preservation standards and best practices is well served by maintaining and promoting collaboration among its internal and external stakeholders. Interdependencies between and among the operations of records producing units of government, legal and statutory requirements, information technology policies and governance, and historical accountability should be systematically addressed.

- ☐ No collaborative digital preservation environment exists in the state/territory.
- ☑ The Archives/RM unit is working to establish a framework for collaborative engagement on digital preservation issues in the state/territory.
- ☑ Under a collaborative digital preservation framework the Archives/RM unit has successfully engaged or currently engaged in identifying specific digital preservation requirements for selected records producing units.
- ☐ Under its collaborative digital preservation framework the Archives/RM unit has successfully engaged or is currently engaged in one or more collaborative digital preservation projects with external stakeholders.
- ☑ Under its collaborative digital preservation framework the Archives/RM unit has successfully engaged in or is currently engaged in identifying the specific digital preservation requirements of most records producing units.
- ☐ The Archives/RM unit continuously monitors and updates the collaborative framework of digital preservation requirements of all records producing units.

5. TECHNICAL EXPERTISE

A critical component in a sustainable digital preservation program is access to professional technical expertise that can proactively address business requirements as well as respond to impacts of evolving technologies. The technical infrastructure and key processes of an ISO 14721 conforming archival repository requires professional expertise in archival storage, digital preservation solutions, and lifecycle electronic records management processes and controls. This technical expertise may exist within the Archives/Records Management unit, be provided by a centralized function or service bureau, or by external service providers and should include an in-depth understanding of critical digital preservation actions and their associated recommended practices.

- ☑ The Archives/RM unit has little or no operational access to specialized professional technical expertise in digital preservation or electronic records management.
- ☐ The Archives/RM unit has operational access to technical expertise (internal or external) that only supports project-based digital preservation initiatives.
- ☐ The Archival/RM unit has operational access to technical expertise (internal or external) in DoD 5015.2 compliant electronic records management software.
- ☐ The Archival/RM unit has operational access to technical expertise (internal or external) that only supports non-conforming ISO 14721 Submission Information Packages (SIP) and Archival Information Packages (AIP).
- ☐ The Archival/RM unit has operational access to technical expertise (internal or external) that supports ISO14721 conforming Submission Information Packages (SIP) and Archival Information Packages (AIP).
- ☐ The Archival/RM unit has operational access to technical expertise that supports all functions of an ISO 14721 conforming archival repository, including long-term digital preservation planning.

6. OPEN STANDARD TECHNOLOGY NEUTRAL FORMATS

A fundamental requisite for a sustainable digital preservation program that ensures long-term access to usable and understandable electronic records is mitigation of obsolescence of file formats. Open standard technology neutral ("OS/TN") file formats are developed in an open, public setting, issued by a certified standards organization, and have few or no technology dependencies. Current preferred OS/TN format examples include:HTML, Plain Text, XML, ODF, and PDF/A for text; CSV for spreadsheets; JPEG 2000 for photographs; PDF/A, PNG, and TIFF for scanned images; SVG for vector graphics; BWF for audio; MPEG-4 and Motion JPEG2000 for video; WARC for web pages. Over time new digital preservation tools and solutions will emerge that will require new OS/TN file formats. OS/TN formats are backwardly compatible so they can support interoperability across technology platforms over an extended period of time.

- ☐ The Archives/RM unit has not adopted any OS/TN file format as a digital preservation format.
- ☑ The Archives/RM unit has adopted at least one OS/TN file format as digital preservation format.
- ☐ The Archives/RM unit has adopted at least three OS/TN file formats as digital preservation formats.
- ☑ The Archives/RM unit has adopted OS/TN for text.
- ☐ The Archives/RM unit has adopted an OS/TN for spreadsheets.
- ☐ The Archives/RM unit has adopted an OS/TN for raster bit map images (scanned and born digital).
- ☐ The Archives/RM unit has adopted an OS/TN for vector graphics.
- ☐ The Archives/RM unit has adopted an OS/TN for audio.
- ☐ The Archives/RM unit has adopted an OS/TN for videos.
- ☐ The Archives/RM unit has adopted an OS/TN for web pages.
- ☐ The Archives/RM unit continuously monitors the sustainability of OS/TN file formats and adopts them as appropriate for use as preservation formats.

7. DESIGNATED COMMUNITY

The organization that has responsibility for preservation and access to permanent legal, operational, fiscal or historical government records is well served through proactive outreach and engagement with its Designated Community. The Archives/Records Management unit has written procedures and formal agreements with records producing units that document the content, rights, and conditions under which the digital archival repository will ingest, preserve, and provide access to electronic records. The Archives/Records Management unit maintains written procedures regarding ingest of electronic records and access to its digital collections. Records Producers will submit fully conforming ISO 14721 Submission Information Packages (SIPs) while Dissemination Information Packages (DIPs) are developed and updated in conjunction with its user communities.

- ☑ There is no written documentation that defines the rights, obligations, and responsibilities of record producing units or designated communities for electronic records held by the archival repository.
- ☑ The Archives/RM unit has informal, ad hoc agreements with selected records producing units that support the transfer of electronic records to the archival repository.
- ☐ The Archives/RM unit periodically analyzes access and use statistics for its archival repositories to identify and address user needs and requirements.
- ☐ The Archives/RM unit has established formal, written agreements that support the transfer of electronic records from selected record producing units.
- ☑ The Archives/RM unit proactively reaches out to selected designated communities to identify their needs and requirements for access to electronic records in the archival repository.
- ☐ The Archives/RM unit works with most records producing units in the state/territory to establish formal written agreements about their rights, obligations, and responsibilities for transferring electronic records to the archival repository.
- ☐ The Archives/RM unit works closely with most designated communities to establish appropriate access priorities that meet their needs and requirements.
- ☐ The Archives/RM unit actively engages with all state/territory records producing units to establish formal written agreements about their rights, obligations and responsibilities for transferring electronic records to the archival repository.
- ☐ The Archives/RM unit works closely with all designated communities to establish appropriate access priorities that meet their needs and requirements.

8. ELECTRONIC RECORDS SURVEY

A trustworthy digital repository cannot fully execute its mission or engage in realistic digital preservation planning without a projected volume and scope of electronic records that will come into its custody. It is likely that some information already exists in approved retention schedules but may require further elaboration as well as periodic updates, especially with regard to preservation ready, near preservation ready, and legacy electronic records held by records producing units.

- ☐ The Archives/RM unit has little or no capability or resources to collect and analyze information about the volume, location, media, format types, and life cycle management requirements for electronic records.
- ☑ The Archives/RM unit relies on existing retention schedules to identify electronic records of permanent historical, fiscal, and legal value in the custody of records producing units.
- ☑ The Archives/RM unit conducts ad hoc, one-time interviews or surveys to identify electronic records of permanent historical, fiscal, and legal value in the custody of selected records producing units.
- ☐ The Archives/RM unit conducts systematic interviews, surveys, and retrospective analysis of existing retention schedules to identify electronic records of permanent historical, fiscal, and legal value in the custody of selected records producing units.
- ☐ The Archives/RM unit periodically analyzes existing retention schedules to identify “at risk” electronic records of permanent historical, fiscal, operational, and legal value in the custody of selected records producing units.
- ☐ The Archives/RM unit supplements retention schedule analysis through collection of information about the volume and location (e.g., shared drives) of “at risk” electronic records of permanent historical, fiscal, operational, and legal value in the custody of records producing units.
- ☑ The Archives/RM unit supplements retention schedule analysis through collection of information about the media and format types of “at risk” electronic records of permanent historical, fiscal, operational, and legal value in the custody of records producing units.
- ☐ The Archives/RM unit has identified preservation-ready and non preservation-ready permanent electronic records in the custody of all of records producing units.
- ☐ The Archives/RM unit uses information (e.g., date eligible for transfer) about electronic records in the custody of all records producing units as an inventory to systematically manage the transfer and ingest of their electronic records.

9. INGEST

A digital archival repository that complies with ISO 14721 has the capability to systematically Ingest(receive and accept) electronic records from records producing units in the form of Submission Information Packages (SIPs), move them to a staging area where virus checks and content and format validations are performed, transform electronic records into designated preservation formats as appropriate, extract metadata from SIPs and write it to Preservation Description Information (PDI), creates Archival Information Packages (AIPs), and transfer the AIPs to the repository’s storage function. This process is considered the minimal workflow for transferring records into a digital archival repository for long-term preservation and access.

- ☑ The Archives/RM unit does not currently accession or ingest electronic records.
- ☐ The Archives/RM unit ingests any available electronic records from records producing units in any format, but assumes only the responsibility to keep the bit stream alive.
- ☐ The archival repository ingests partially conforming ISO 14721 SIPs.
- ☐ Partially conforming ISO 14721 SIPs are ingested and held in a staging area while virus checks and format validations are manually executed.
- ☐ Partially conforming ISO 14721 AIPs are manually transferred from a staging area to archival storage.
- ☐ Fully conforming ISO 14721 SIPs are ingested and checked for virus and format validations with semi-automated tools.
- ☐ Fully conforming ISO 14721 AIPs are produced and transferred to archival storage through semi-automated tools.
- ☐ Fully conforming ISO 14721 SIPs are ingested and automatically checked for virus and format validations.
- ☐ Fully conforming ISO 14721 AIPs are automatically produced and transferred to archival storage.

10. ARCHIVAL STORAGE

ISO 14721 delineates systematic automated storage services that support receipt and validation of successful transfer of Archival Information Packages (AIPs) from ingest, creation of Preservation Description Information (PDI) for each AIP that confirms its "fixity" during any preservation actions through the capture and maintenance of error logs, updates to PDI, including transformation of electronic records to new formats, production of Dissemination Information Packages (DIPs) for Access, and collection of operational statistics.

- ☑ The Archives/RM unit only has access to primitive non-conforming archival storage (e.g., CDs/DVDs).
- ☐ One storage tier level is used in the archival repository for partially conforming ISO14721 AIPs.
- ☐ A single instance of an archival storage repository is used for the storage of partially conforming ISO 14721 AIPs.
- ☐ Two or more storage tier levels are used to store partially conforming ISO 14721 AIPs.
- ☐ Two geographically separated instances of an archival storage repository are used for the storage of partially conforming ISO 14721 AIPs.
- ☐ Manual capture of partially conforming Preservation Description Information (PDI) establishes the provenance of partially conforming ISO 14721 AIPs.
- ☐ Manual capture of selected partially conforming Archival Storage operational statistics to support ad hoc digital preservation planning.
- ☐ Two or more storage tier levels are used to store conforming ISO 14721 AIPs.
- ☐ Two geographically separated instances of an archival storage repository are used for the storage of conforming ISO 14721 AIPs.
- ☐ Conforming ISO 14721 Preservation Description Information (PDI) that establishes the provenance of compliant AIPs (e.g., format transformations, device/media renewal, integrity checks, and access/rights protection) is captured with semi-automated tools.
- ☐ Semi-automated capture of conforming ISO 14721 archival storage operational statistics that support systematic digital preservation planning.
- ☐ Automated capture of Preservation Description Information (PDI) establishes the provenance of conforming ISO 14721 AIPs.
- ☐ Automated capture of conforming archival storage operational statistics supports on- going comprehensive digital preservation planning.
- ☐ Three geographically separated instances of an archival storage repository are used for the storage of conforming ISO 14721 AIPs.

11. DEVICE/MEDIA RENEWAL

No known digital device or storage medium is invulnerable to decay and obsolescence. A foundational digital preservation capability is ensuring the readability of the bitstreams underlying the electronic records. ISO 14721 specifies that a trustworthy digital repository's storage devices and storage media should be monitored and renewed (" refreshed") periodically to ensure that the bit streams remain readable over time. A projected life expectancy of removable storage media does not necessarily apply in a specific instance of storage media. Hence, it is important that a trustworthy digital repository have a protocol for continuously monitoring removable storage media (e.g., magnetic tape, external tape drive, or other media) to identify any that face imminent catastrophic loss. Ideally, this renewal protocol would automatically execute renewal after review by the digital archival repository.

- ☑ The Archives/RM unit has no device/media renewal protocol in place.
- ☐ Current practice mandates archival repository device/media renewal when they are on the verge of becoming obsolescent.
- ☐ Current practice mandates archival repository device/media renewal on a regularly scheduled basis (e.g., every ten years).
- ☐ An annual device/media inspection program identifies archival repository device/storage media that face imminent catastrophic data loss.
- ☐ The archival repository’s device and media renewal program continuously monitors the potential loss of the readability of electronic records and automatically replaces devices/storage media and writes the records to new storage media.

12. INTEGRITY

A key capability in ISO 14721 conforming digital repositories is ensuring the integrity of the records in its custody, which involves two related preservation actions. The first action generates a cryptographic hash algorithm which takes any digital object regardless of size or content type and normalizes it to a fixed length bit stream (e.g., 128 bits). This fixed length bit stream is called a hash digest and it serves as a digital fingerprint of a larger digital object. The second action involves integrity fixity that supports an unbroken electronic chain of custody captured in Preservation Description Information (PDI) in AIPs.

- ☑ The archival repository has no documented procedure for integrity protection of electronic records in its custody.
- ☐ The archival repository generates and preserves MD-5 hash digests before and after device/media renewal and other archival storage preservation actions.
- ☐ The archival repository generates and preserves SHA-1 hash digests before and after device/media renewal and other internal preservation actions for partially conforming ISO 14721 AIPs.
- ☐ The archival repository generates SHA-2 hash digests before and after device/media renewal and other internal preservation actions for all fully conforming ISO 14721 AIPs and stores them in the Preservation Description Information (PDI) of the AIPs.
- ☐ The archival repository encapsulates fully conforming ISO 14721 AIPs in XML and signs them with a digital signature.
- ☐ Integrity protection procedures are continuously evaluated and updated as new tools and approaches become available.

13. SECURITY

Contemporary enterprise-wide information systems typically execute a number of shared or common services that may include inter-process communication, name services, temporary storage allocation, exception handling, role based access rights, security, backup and business continuity, and directory services, among others. An ISO 14721 conforming archival repository is likely to be part of an information system that may routinely provide some or perhaps all of the core security, backup, and business continuity services including firewalls, role based access rights, data transfer integrity validations, logs for all preservation activities, including failures and anomalies to demonstrate an unbroken chain of custody.

- ☑ Currently, the archival repository does not have formal disaster recovery, backups, or firewall procedures in place to protect the security of electronic records.
- ☐ The security of electronic records in the archival digital repository is protected through disaster recovery procedures.
- ☐ The security of electronic records in the archival digital repository is protected through comprehensive firewall protection.
- ☐ The security of electronic records in the archival digital repository is protected through comprehensive role based access rights management.
- ☐ The archival repository continuously monitors security protection processes and revises them in response to evolving technology capabilities and changing business requirements

14. PRESERVATION METADATA

A digital archival repository collects and maintains metadata that describes actions associated with custody of permanent records including an audit trail that documents preservation actions carried out, why and when they were performed, how they were carried out and with what results. A current best practice is the use of a PREMIS-based Data Dictionary to support an electronic chain of custody that documents authenticity over time as preservation actions are executed. Capture of all related metadata, transfer of the metadata to any new formats/systems, and secure storage of metadata are critical. All metadata is stored in the Preservation Description Information (PDI) component of conforming AIPs.

- ☑ Little or no preservation metadata is created, captured or maintained for electronic records.
- ☐ Minimal preservation metadata is available to support electronic records in the archival repository.
- ☐ A partially conforming ISO 14721 preservation metadata scheme (like Dublin Core) is in place for electronic records in the archival repository.
- ☐ A preservation metadata schema is in place that includes a data dictionary that identifies ISO 14721 conforming comprehensive preservation metadata specifications.
- ☐ A PREMIS-based preservation metadata schema is in place that is continuously reviewed and updated to take into account new and different types electronic records that are transferred to an archival repository.

15. ACCESS

Organizations with a mandate to support public access to permanent government records are subject to authorized restrictions. An ISO 14721 conforming archival repository will provide consumers with trustworthy records in “disclosure free” Dissemination Information Packages (DIPs) redacted to protect, privacy, confidentiality, and other rights where appropriate, and searchable metadata that users can query to identify and retrieve records of interest to them. Production of DIPs is tracked, especially when they involve extractions, to verify their trustworthiness and to identify query trends that are used to update electronic accessibility tools to support these trends.

- ☑ The archival repository has no capability to support access to electronic records in its custody.
- ☐ The archival repository provides copies of partially conforming ISO 14721 DIPs in at least one (1) format (e.g., digital photographs in JPEG 2000).
- ☐ The archival repository supports access only to partially conforming ISO 14721 DIPs in two (2) open standard technology neutral (OS/TN) formats (e.g., PDF/A, JPEG 2000, or TIFF).
- ☐ The ISO 14721 conforming archival repository produces DIPs in at least six (6) open standard technology neutral (OS/TN) formats.
- ☐ The conforming ISO 14721 archival repository analyzes user query trends to identify the need for updated accessibility tools.
- ☐ The conforming ISO 14721 archival repository disseminates DIPs containing records in any format that users request.
- ☐ The conforming ISO 14721 archival repository enables redaction of electronic records with access restrictions in its custody where appropriate.

BIBLIOGRAPHY

[1] Filezilla - the free ftp solution. https://filezilla-project.org/. Accessed: 2014-03-31.

[2] Free ftp client, secure file transfer software. http://www.coreftp.com/. Accessed: 2014-03-31.

[3] Interpares 2 terminology database. http://www.interpares.org/ip2/ip2_terminology_db.cfm/. Accessed: 2014-03-21.

[4] Libraryofcongress/bagit-java github. https://github.com/LibraryOfCongress/bagit-java. Accessed: 2014-03-30.

[5] Metadata encoding transmission standard. http://www.loc.gov/standards/mets/. Accessed: 2014-03-24.

[6] Smartftp - ftp client. http://www.smartftp.com/. Accessed: 2014-03-31.

[7] Concepts, principles, and methods for the management of electronic records. *Information Society*, 14(4):271, 2001.

[8] Jan Askhoj, Mitsuharu Nagamori, and Shigeo Sugimoto. Archiving as a service: a model for the provision of shared archiving services using cloud computing. In *Proceedings of the 2011 iConference*, iConference '11, pages 151–158, New York, NY, USA, 2011. ACM.

[9] VJJM Bekkers. New forms of steering and the ambivalency of transparency. *Public administration in an information age*, pages 341–357, 1998.

[10] Heiko Bock. *The Definitive Guide to NetBeans Platform 7*. Apress, 2011.

[11] John Brooke. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189:194, 1996.

[12] Ben Collins-Sussman, Brian Fitzpatrick, and Michael Pilato. *Version control with subversion*. O'Reilly, 2004.

[13] Oxford English Dictionary. Oxford english dictionary. 2006.

[14] Luciana Duranti and Kenneth Thibodeau. The concept of record in interactive, experiential and dynamic environments: the view of interpares*. *Archival Science*, 6(1):13–68, 2006.

[15] Marilyn Geller. Submission information package (sip) specification. *Harvard E-Journal Archive, Harvard University Library Office for Information Systems*, 2001.

[16] David Giaretta. *Advanced digital preservation*. Springer, 2011.

[17] Yunhong Gu, Robert Grossman, Xinwei Hong, and Marco Mazzucco. Using udp for reliable data transfer over high bandwidth-delay product networks. *Laboratory for Advanced Computing, University of Illinois at Chicago*, 2003.

[18] Yunhong Gu, Xinwei Hong, Marco Mazzucco, and RL Grossman. Sabul: A high performance data transfer protocol. *submitted to IEEE Communications Letters*, 2003.

[19] Andrew Hanushevsky, Artem Trunov, and Les Cottrell. Peer-to-peer computing for secure high performance data copying. In *In Proc. of the 2001 Int. Conf. on Computing in High Energy and Nuclear Physics (CHEP 2001), Beijng*. Citeseer, 2001.

[20] Eric He, Jason Leigh, Oliver Yu, and Thomas A DeFanti. Reliable blast udp: Predictable high performance bulk data transfer. In *Cluster Computing, 2002. Proceedings. 2002 IEEE International Conference on*, pages 317–324. IEEE, 2002.

[21] Jansen. Digital records forensics: Ensuring authenticity and trustworthiness of evidence over time. *University of British Columbia, InterPARES Trust Research Team*, 2010.

[22] Jansen. Authenticity in records systems: Emerging research in digital preservation. *University of British Columbia, InterPARES Trust Research Team*, 2014.

[23] Duranti Jansen. Authenticity of digital records: An archival diplomatics framework for digital forensics. *University of British Columbia, InterPARES Trust Research Team*, 2011.

[24] Jon Loeliger and Matthew McCullough. *Version Control with Git: Powerful tools and techniques for collaborative software development*. O'Reilly Media, Inc., 2012.

[25] Mark Meiss. Tsunami: A high-speed rate-controlled protocol for file transfer. *2004-09-28). http://Steinbeck. ucs. indiana. edu/-mmeiss/papers/tsunami. pdf*, 2004.

[26] Reagan Moore, Chaitan Baru, Arcot Rajasekar, Bertram Ludaescher, Richard Marciano, Michael Wan, Wayne Schroeder, and Amarnath Gupta. Collection-based persistent digital archives-part 1. *D-Lib magazine*, 6(3):1082–9873, 2000.

[27] NASA. The apollo 11 telemetry data recordings: A final report. *NASA Center for AeroSpace Information (CASI)*, 2010.

[28] Quyen L Nguyen and Dyung Le. Archival asset package design concept for an oais system. In *Proceedings of the 2010 Roadmap for Digital Preservation Interoperability Framework Workshop*, page 4. ACM, 2010.

[29] Eun G Park. Understanding" authenticity" in records and information management: Analyzing practitioner constructs. *American Archivist*, 64(2):270–291, 2001.

[30] Amy Rudersdorf, Dean Farrell, and Lisa Gregory. Electronic records processing: it's a cinch! In *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries*, pages 375–376. ACM, 2012.

[31] Harimath Sivakumar, Stuart Bailey, and Robert L Grossman. Psockets: The case for application-level network striping for data intensive applications using high speed wide area networks. In *Proceedings of the 2000 ACM/IEEE conference on Supercomputing (CDROM)*, page 37. IEEE Computer Society, 2000.

[32] MacKenzie Smith, Mary Barton, Mick Bass, Margret Branschofsky, Greg McClellan, Dave Stuve, Robert Tansley, and Julie Harford Walker. Dspace: An open source dynamic digital repository. 2003.