

**Universidad de las Ciencias Informáticas**

**Facultad 2**



**Sistema para la preservación digital a largo plazo de la información  
científico-técnica de la Biblioteca de la UCI**

Trabajo de Diploma para optar por el título de Ingeniero en Ciencias  
Informáticas

**Autores:**

Yaneisi Mayea Toledo

Alfredo Viset Curiaut

**Tutor:**

MSc. Edisnel Carrazana Castro

**Cotutor:**

Ing. Osmaily Durán Pérez

**La Habana, junio de 2016.**

**“Año 58 de la Revolución”**

*"La sabiduría de una persona no se cataloga en toda la teoría que sabe, sino en su forma de poner en práctica su conocimiento y vivirlo."*

*Enrique Monterrosa*



## **Declaración de autoría**

Declaramos que somos los únicos autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmo la presente a los \_\_\_\_\_ días del mes \_\_\_\_\_ del año \_\_\_\_\_.

---

Yaneisi Mayea Toledo.

---

Alfredo Viset Curiaut.

---

Edisnel Carrazana Castro.

---

Osmaily Durán Pérez.

## *Dedicatoria*

*A mis padres Edel y Yadira, por su amor incondicional en todo tiempo, porque a lo largo de los años han sido mi apoyo y mi principal motivación para llegar al final de mi carrera universitaria.*

*A María y Rafael mis abuelos y mi tío Félix, que me apoyaron en todo lo que he emprendido en mi vida.*

*Yaneís*

*Al Señor Jesús, por cuidar de mí todos estos años y nunca dejarme solo, su fidelidad ha sido grande.*

*A la mejor madre del mundo, mi mamá. Por su entrega incondicional, su amor, sus instrucciones y enseñanzas por ello hoy estoy aquí.*

*Al varón, por ser el padre que es. Por guiarme, enseñarme y ser un ejemplo para mí.*

*A mis hermanos Víctor y Edel.*

*Alfredo*

## *Agradecimientos*

*A Dios primero que todo porque siempre ha sido mi amigo fiel, mi fortaleza y él le ha dado sentido a mi vida.*

*A mis padres Yadira y Edel por siempre animarme y suplirme de todo cuanto he necesitado.*

*A mis abuelos María, Rafael que siempre me han tenido dentro de sus prioridades, me han apoyado en los momentos buenos y malos, siempre orando por mí.*

*A mis tíos Félix y Osmany que no importaba cuál fuera la circunstancia siempre estaban dispuestos a ayudarme y siempre me animaron. En fin, a toda mi familia en general abuelos, primos y tíos.*

*A los mejores amigos que siempre tendré, los que siempre han estado conmigo y me han ayudado en todos estos años: Juan (Querído), Yasel, Adyana (apileta), Yerandy, Yasmary (kiriña) y Danneris (amiguíis).*

*A los mejores compañeros y amigos de todos los días: Yusnelys (Buenagon mi compañera de cuarto en los últimos años) y Barbaro.*

*A mis hermanos en Cristo que siempre me apoyaron en oración y con los que he compartido incontables momentos buenos y malos, entre ellos. Todos son especiales, pero no quería dejar de mencionar a Yoandry, Mariam, Lexy y a Yisel.*

*A mis tutores Edisnel y Osmaíly que nos ayudaron en la realización de la tesis dándonos aliento, ánimo y orientación. Sobre todo, por su paciencia y el empeño que pusieron para que todo saliera bien.*

*A mi amigo y compañero de tesis Alfredo, que durante todo el tiempo que hemos compartido siempre me ha ayudado y soportado. Le agradezco porque también es gracias a él que hoy estamos aquí y por todos estos años.*

*A la oponente que con sus sugerencias fue posible un mejor resultado, al tribunal y a todos los presentes aquí hoy. Gracias por estar aquí y compartir este día tan especial con nosotros. **Yaneísí***

## *Agradecimientos*

*Al que me guarda de todo mal, al que me guía en cada momento de mi vida, al que me cuidó aún desde el vientre de mi madre, al que prometió estar conmigo todos los días hasta el fin del mundo, a aquel que murió para darme vida, al que me limpió de todos mis pecados, al Santo, al Fiel, al Verdadero, el que abre y ninguno cierra, cierra y nadie puede abrir, aquel que es la principal razón de mi vida... a Jesús. Todo se lo debo a Él.*

*A mis padres Víctor Manuel y María del Carmen por su cuidado, por haberme dado todo lo que tienen, porque en los 25 años de vida que tengo nunca se les acabó el amor, nunca se les acabaron las palabras de consuelo, de ánimo ni de sabiduría. Porque nunca se le acabaron los ánimos para ir a verme un domingo a la escuela, aunque estuvieran enfermos. Porque en el momento que salí de esta universidad supieron recibirme, y darme palabras de ánimo, de reprensión y de perseverancia... y mi mamá decirme una frase que jamás he olvidado "lo que uno empieza lo termina". A ellos gracias.*

*A mis hermanos Edel y Víctor por ir siempre delante de mí en las escuelas. Por retarme a mantener en alto los "Viset Curiaut". Por todas sus palabras, y a vítico por su ejemplo.*

*A mi tía Maricel por ayudarme cuando lo necesité y a toda mi familia.*

*A mis hermanos Edísnel y Osmaílí por todo su tiempo y esfuerzo dedicado a esta tesis, sin su ayuda este trabajo no habría tenido un final feliz. No tengo como agradecer la ayuda que nos han dado, solo darles las gracias y que el Señor se los recompense y nunca la falten la gracia y la bendición de Dios.*

*A personas que han hecho que mi vida aquí sea mejor: al entrenador del equipo americano(Joel), a trinidad(Yunexy), a mi hermano y compañero de equipo de atletismo(Ge), a mis primeros compañeros de casa Juan, Sergio, Jose, la Rubia entre otros. A los amigos, que han estado en las buenas y en las malas: la india de Pinar (Buenagon), a Barbaro, Niharby, Ramiro.*

*Un agradecimiento grande a todos mis hermanos de la UCI, una de las mejores cosas que me han pasado son los tiempos que hemos compartidos juntos. Sin ellos los años para mí fueran siglos y los días meses, a ellos gracias. En especial a Eliany, Yohandry, Alayo, Jany, Adelaida, Juan, Adyana y Eliezer.*

*Y por último agradecer a la sspirituana por todo su tiempo, entrega y dedicación. Por preocuparse por mí en todo momento, cuando estaba enfermo o cuando iba de competencia. Por cuidar de mí como si fuera mi madre, encontré una buena amiga. Gracias también a su mamá, su papá y a su familia por todo lo que han hecho. Por tratarme bien cuando estuve por allá por su tierra. Muchas gracias de verdad que el Señor los bendiga por siempre. **Alfredo***

### Resumen

En la presente investigación se estudian los principales elementos conceptuales de la preservación digital, así como las vulnerabilidades y amenazas a las que se enfrentan los documentos en formato digital. Se describen las estrategias y los sistemas más utilizados para garantizar la preservación a largo plazo. Además, se especifica el lenguaje, las herramientas y las tecnologías utilizadas para la implementación de la propuesta de solución. Se realiza una caracterización del sistema de preservación digital a implantar, para ello se describen sus principales funcionalidades y se hace un análisis de su arquitectura, identificándose los elementos que deben personalizarse para su adaptación a las necesidades de la Biblioteca de la UCI. Finalmente se aplica el método para validar la propuesta de solución.

#### **Palabras claves:**

Documentos digitales, estrategias de preservación, preservación digital, sistemas de preservación digital.



## Índice

Introducción.....	1
Antecedentes de la investigación:.....	1
Situación problemática.....	3
Capítulo 1: Aspectos teóricos sobre preservación digital.....	6
1.1 Preservación Digital.....	6
1.2 Entorno Digital.....	7
1.3 Documento digital.....	8
1.4 Formatos.....	10
1.5 Modelo OAIS.....	12
1.6 Vulnerabilidades y amenazas a la preservación digital.....	12
1.7 Estrategias de preservación digital.....	16
1.8 Descripción general de los sistemas de preservación.....	18
1.9 Selección del Sistema de Preservación.....	20
1.10 Lenguaje, herramientas y tecnologías a utilizar.....	22
1.11 Metodología a utilizar.....	23
Capítulo 2: Propuesta de solución.....	25
2.1 Evaluación de Cambios.....	25
2.2 Desarrollo de aplicaciones.....	26
2.2.1 Arquitectura de Archivematica.....	26
2.2.2 Patrón de diseño de Archivematica.....	31
2.2.3 Funcionalidades de Archivematica.....	31
2.2.4 Personalización de la interfaz.....	40
2.2.5 Migración de datos.....	41
2.3 Gestión de liberaciones y despliegue.....	45
2.3.1 Instalación y configuración de Archivematica en el entorno real.....	45
2.3.2 Capacitación.....	46

2.4 Validación y Pruebas de Servicios .....	47
2.4.1 Pruebas de aceptación.....	47
2.5 Beneficios de la propuesta.....	49
Capítulo 3: Validación de la propuesta de solución .....	50
3.1 Método Delphi.....	50
3.2 Fase 1: Formulación del problema.....	50
3.3 Fase 2: Elección de expertos.....	50
3.4 Fase 3: Elaboración y lanzamiento de los cuestionarios .....	54
3.5 Fase 4: Desarrollo práctico y explotación de resultados.....	54
3.6 Establecimiento de la concordancia entre los Expertos.....	56
Conclusiones generales .....	59
Recomendaciones.....	60
Referencias Bibliográficas .....	61
Anexo 1 .....	65
Anexo 2.....	66
Anexo 2.....	68
Anexo 4.....	69
Anexo 5.....	70
Anexo 6.....	71
Anexo 7.....	72
Anexo 8.....	73

## **Introducción**

En la actualidad las tecnologías, los medios y dispositivos tecnológicos se desarrollan cada vez más, esto ha propiciado que, en conjunto con ellos, aumente el volumen de información en formato digital. Con la información en formato digital se obtienen muchos beneficios, ya que evita el deterioro, permite que esta sea más accesible y pueda ser consultada por diferentes usuarios al mismo tiempo. Pero que se encuentre en formato digital no garantiza su existencia a largo plazo, ya que está sujeta a amenazas como el deterioro de los medios de almacenamiento, ya sean CD-ROM, discos duros u otros medios de almacenamiento, el borrado de la información de manera accidental o intencionada por parte usuarios y operadores, al igual que el acceso no autorizado a la información. Hoy en día quedan obsoletos a una mayor velocidad los medios tecnológicos, los formatos y el propio software, producto al desarrollo constante de los mismos; esto implica que, aun siendo capaces de guardar el fichero por un largo tiempo, no se pueda garantizar el acceso a su información.

La Universidad de las Ciencias Informáticas (UCI), desde su creación ha sido de impacto en la sociedad cubana y en otros países, por los softwares desarrollados para la informatización de diversos sectores como la economía y la salud. En la UCI, las actividades fundamentales están enfocadas principalmente a la docencia, la producción y la investigación, generándose en cada una un gran cúmulo de importante información digital, en diferentes formatos. Esta información juega un papel fundamental en las rendiciones de cuenta y la toma de decisiones.

Como parte importante de la documentación generada está la información científico-técnica, la cual tiene como fuentes fundamentales: trabajos de diploma, tesis de maestría y doctorado, artículos científicos publicados en revistas, así como trabajos presentados en eventos. Esto provocó que la dirección de la Universidad creara proyectos de informatización para la gestión de la información científico-técnica, de lo cual se encarga el Centro de Información Científico-Técnica, específicamente la Biblioteca como parte del mismo.

### **Antecedentes de la investigación:**

En dos trabajos de diploma de la UCI se trata el tema de la conservación de archivos históricos (Vazquez, 2013) (Blanco, 2011), analizándose la conservación en formato duro y no en el digital, por lo que no constituyen trabajos de relevancia para la presente investigación. Es por eso que el tema preservación digital es novedoso en la Universidad.

Además, se desarrolló la tesis de maestría Propuesta de requisitos funcionales para la preservación digital a largo plazo de la información científico-técnica de la Universidad de las Ciencias Informáticas, desarrollada en la Universidad de la Habana, en la facultad de Ciencias de la Información (Carrazana, 2015), la cual puede resultar muy útil el análisis del contexto en que se realiza esta investigación dado que se enmarca en la información científico-técnica de la UCI.

A nivel nacional se han desarrollado investigaciones de gran importancia sobre la preservación digital, como la tesis doctoral Sistema de Gestión Integral de Documentos de archivo para empresas de la construcción del territorio de Camagüey (Campillo, 2010), se destaca también por su alcance, la tesis doctoral Propuesta de requisitos funcionales para la gestión de documentos archivísticos electrónicos en la administración central del estado cubano (Mena Mugica, 2006).

El plano internacional constituye la fuente más importante de información, pues existen varios proyectos y estándares que pueden contribuir al desarrollo del presente trabajo.

Tal es el caso del proyecto Interpares (*International Research on Permanent Authentic Records in Electronic Systems*), surgido en 1999, el cual tiene como objetivo desarrollar los conocimientos esenciales para la preservación a largo plazo de documentos auténticos creados y/o mantenidos en forma digital y proporcionar la base para las normas, políticas, estrategias y planes de acción capaces de asegurar la longevidad de este tipo de material y la capacidad de los usuarios para confiar en su autenticidad (InterPARES., 2013). Aún se utilizan las teorías y metodologías para la preservación de documentos electrónicos planteados por el proyecto.

El modelo de referencia OAIS (*Reference Model for an Open Archival Information System*) proporciona un marco para la comprensión y el mayor conocimiento de los conceptos archivísticos necesarios para la conservación y el acceso de la información digital a largo plazo (CCSDS, 2012). Provee un conjunto de estándares para el desarrollo de sistemas electrónicos para la preservación a largo plazo.

La Biblioteca de la UCI tiene entre sus funciones la gestión de la información científico-técnica de la Universidad. Para ello cuenta con un Repositorio Institucional basado en el software de código abierto Dspace, en su versión 5.4. “Este es un sistema gratuito, libre, fácil de instalar, altamente configurable para poder ser adaptado a cualquier institución y permite el acceso abierto a todo tipo de contenido digital (textos, imágenes, etc)”. (Dspace, 2016)

El repositorio tiene dos módulos fundamentales, uno para la administración, que es el módulo principal y otro para la búsqueda. Dentro de las funciones del primer módulo están: permitir el control del sistema, el trabajo con las comunidades y colecciones, así como la administración de usuarios y grupos, flujos

de trabajo, estadísticas y otros. A este módulo solo pueden acceder los usuarios con permisos de administración. El segundo módulo presenta la búsqueda como un aspecto esencial para el repositorio, esta se puede realizar de múltiples maneras: por referencia externa utilizando *handles* (manejadores), por palabras claves en los metadatos o extraídas de texto completo, por título, autor, fecha, materia, etc.

El sistema permite la descripción de los recursos digitales a través de metadatos descriptivos haciendo uso del formato Dublin Core. Este último puede ser cambiado por otro o modificado a través del módulo de Administración.

### Situación problemática

El Repositorio Institucional trata algunas de las amenazas a las que está expuesta la información digital, dado que realiza salvallas periódicas de los datos, lo que le permitiría restaurarlos en caso de pérdidas. Sin embargo, a pesar de sus ventajas para la gestión de la información, según una investigación realizada por (Carrazana, 2015), la versión de Dspace utilizada por el repositorio de la Universidad también presenta algunas limitaciones que le impiden garantizar la preservación de la información digital a largo plazo:

- No garantiza la integridad de los documentos durante su ingreso y para el envío de la información no se utiliza un canal de transferencia seguro.
- No garantiza la autenticidad de los documentos digitales durante el ingreso, pues no utiliza la firma digital o el sellado de tiempo, para verificar que los documentos fueron creados confiables, es decir, por quien dice ser el autor y en el momento en que se afirma fueron creados.
- No garantiza la autenticidad de los documentos durante el acceso.
- No implementa funcionalidades asociadas a la Auditoría la cual es de vital importancia para detectar y prevenir fallos en el sistema provocados por las vulnerabilidades y amenazas.
- No implementa funcionalidades asociadas a la Migración, lo cual excluye la presencia de mecanismos para asegurar que los documentos digitales puedan ser visualizados por las aplicaciones existentes en el futuro, aún si estos se conservan intactos en los medios de almacenamiento.
- No implementa funcionalidades para mantener una conexión segura y sin ambigüedades entre los metadatos y los documentos digitales que describen, debido a lo cual no se puede asegurar que los documentos son auto-documentados a través del tiempo.
- No registra metadatos adecuados para la descripción técnica.

Considerando que la información científico-técnica en formato digital que se genera en la UCI tiene gran valor desde el punto de vista académico y científico, y que las plataformas informáticas cambian continuamente dejando obsoletos los medios de almacenamiento, software y formatos, resulta de gran importancia realizar acciones encaminadas a reducir las pérdidas a las que puede estar sometido el patrimonio intelectual de la Universidad.

Por lo anteriormente planteado surge como **problema a resolver**: el Repositorio Institucional de la Biblioteca de la UCI no implementa las funcionalidades necesarias para realizar la preservación digital, por lo que no garantiza a largo plazo la disponibilidad y autenticidad de la información científico-técnica. Con el siguiente **objeto de estudio**: la preservación digital a largo plazo de la información científico-técnica. Para dar solución al problema planteado se tiene el siguiente **Objetivo general**: Implantar un sistema de preservación digital en la Biblioteca de la UCI, para garantizar la disponibilidad y autenticidad de la información a largo plazo. Enmarcada la investigación en el **campo de acción**: los sistemas de preservación digital.

Para el desarrollo de la investigación fue necesario trazar las siguientes **tareas de la investigación**:

1. Descripción de la naturaleza de la información en formato digital y los aspectos teórico-conceptuales para una mejor comprensión de la investigación.
2. Caracterización de los diferentes sistemas de preservación digital a largo plazo para posteriormente hacer la elección de uno de ellos como alternativa de solución al problema planteado.
3. Instalación y configuración del sistema de preservación digital seleccionado para su despliegue en la Biblioteca de la UCI.
4. Internacionalización del sistema de preservación digital al español para un mayor entendimiento por parte de los administradores que harán uso del software.
5. Migración de los documentos digitales gestionados por el Repositorio Institucional para su almacenamiento en el sistema de preservación digital.
6. Capacitación sobre el sistema de preservación digital a los especialistas de la Biblioteca encargados de su utilización para que sea correctamente manipulado.
7. Aplicación del método Delphi para validar que la propuesta permite realizar la preservación a largo plazo de la información científico-técnica.

Para el desarrollo de esta investigación y el logro de su objetivo, se utilizaron los siguientes métodos:

### Métodos teóricos

Análítico-Sintético: permitió hacer un análisis de los principales aspectos teóricos relacionados con la investigación y llegar a conclusiones de los mismos.

Histórico-Lógico: permitió estudiar los sistemas de preservación existentes y de esta forma conocer sus principales características y cómo funcionan.

### **Métodos empíricos**

Análisis documental: permitió establecer un marco teórico conceptual y estudiar el estado del conocimiento y de las prácticas relativas a la preservación a largo plazo en ambiente digital a nivel internacional y nacional.

Encuesta: se le presentaron preguntas en formato digital a los encuestados, para seleccionar los expertos del método Delphi y para evaluar el sistema de preservación.

El trabajo de diploma está estructurado en introducción, tres capítulos, conclusiones, recomendaciones, referencias bibliográficas y anexos. Los capítulos se desglosarán de la siguiente forma:

**Capítulo 1:** Aspectos teóricos de la preservación digital. Se estudian los aspectos conceptuales necesarios para el análisis de la naturaleza de la información en ambiente digital, las amenazas fundamentales a las que está expuesta, así como los sistemas y modelos fundamentales para su preservación a largo plazo. Se especifica el lenguaje, las herramientas y las tecnologías utilizadas para la implementación de la propuesta de solución.

**Capítulo 2:** Propuesta de solución. Se realiza una caracterización del sistema de preservación digital a implantar, para ello se describen sus principales funcionalidades y se realiza el análisis de su arquitectura. Además, se describe la implantación del sistema en la Biblioteca de la UCI, para lo cual se identifican los elementos que deben personalizarse para su adaptación al contexto. Finalmente, se realizan pruebas de aceptación al sistema para verificar que se ajusta a las expectativas del usuario final.

**Capítulo 3:** Validación de la propuesta de solución. Se aplica el método Delphi para validar que el sistema implantado posee las funcionalidades requeridas para la preservación digital.

## **Capítulo 1: Aspectos teóricos sobre preservación digital**

En el presente capítulo se realizó un estudio sobre preservación digital, abordándose varios elementos teóricos como cualidades, vulnerabilidades y amenazas de los documentos digitales. Se llevó a cabo un análisis detallado sobre los sistemas de preservación más usados en el mundo. Además, se especificaron las herramientas que se seleccionaron para dar solución al problema en cuestión.

### **1.1 Preservación Digital**

Algunos de los conceptos importantes revisados durante la investigación sobre la preservación digital a largo plazo son:

- ❖ “Puede definirse como el conjunto de actuaciones relativas a la preservación de los documentos existentes en formato digital. Así, si la entendemos como un conjunto de actividades, entre las que se incluye la conservación, y que están destinadas a que un objeto perdure el mayor tiempo posible en su estado original, la preservación digital supondría llevar a cabo estas tareas sobre información en formato digital, preocupándonos no solo por el mantenimiento del objeto, sino también (y principalmente) por su contenido informativo.”

(Rothenberg, 1999)

- ❖ “La preservación consiste en mantener la capacidad de presentar los elementos esenciales de objetos digitales auténticos.”

(Biblioteca Nacional de Australia, 2003)

- ❖ Conseguir que los documentos sean y permanezcan “auténticos, completos, accesibles y comprensibles, procesables, y potencialmente reusables”, indica que la preservación correcta estará asegurada, entre otras cuestiones si: “los metadatos indican el nombre y la versión del formato de cada archivo y la identificación del software con que ha sido creado y con que se ha realizado la última modificación”.

(Internacional council on archives, 2010)

- ❖ “Son las acciones necesarias para asegurar un acceso continuado y pertinente a la información digital durante el tiempo que sea requerido y para cualquier finalidad legítima.”

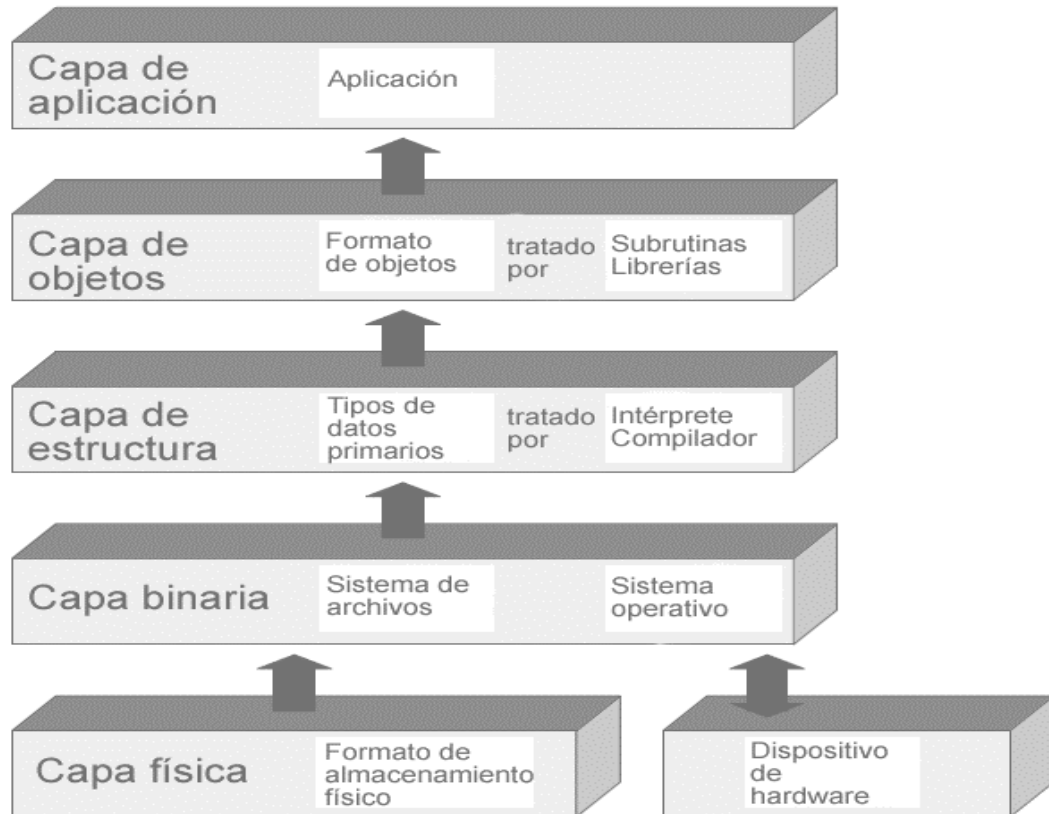
(Keefer A, 2014)

De los conceptos analizados se puede concluir que la preservación va más allá de almacenar y conservar la secuencia de bits de un documento digital por un largo periodo de tiempo. La preservación consiste en lograr que la información pueda estar accesible y que mantenga sus propiedades de autenticidad y disponibilidad.



### 1.2 Entorno Digital

Hoy en día el entorno en que es creada la información aporta características a la misma, que deben ser preservadas de igual manera. Es fundamental tener en cuenta este hecho, dado que la obsolescencia tecnológica es una de las principales amenazas a la que está sometida la información digital. La siguiente figura representa en forma de capas los principales elementos que intervienen e interactúan en la creación, codificación, almacenamiento y visualización de los documentos digitales.



**Figura 1: Modelo de información en capas.** Adaptada de (Lupovici, C., & Masanès, J, 2000).

Las capas representadas en esta figura muestran las diversas formas en que son configurados cada uno de los documentos digitales para ser leídos y entendidos por el usuario. Cada una de estas capas se describen de la siguiente manera (Lupovici, C., & Masanès, J, 2000):

1. En la capa física un documento digital es ante todo presentado en un medio físico o de comunicación, a través de un formato de almacenamiento físico, generalmente estandarizado como por ejemplo ISO 9660 para el CD-ROM. En esta capa, la información es representada como cadenas de bits en un medio físico como dominios magnéticos o como voltajes. Los controladores de dispositivos y chips integrados en la interfaz de almacenamiento físico gestionan este formato a fin de entregar el flujo de bits a la siguiente capa.

2. En la capa binaria el flujo de bits es organizado en bloques etiquetados en una forma independiente del medio. El sistema operativo, el cual gestiona el sistema de archivos, proporciona la funcionalidad de este nivel.
3. La capa de estructura convierte los flujos de bits/byte que recibe de la interfaz de la capa binaria en estructuras direccionables de tipos de datos primitivos, para ser manipulados por los programas. En los sistemas computacionales modernos los compiladores e intérpretes de los lenguajes de programación proporcionan la funcionalidad de este nivel.
4. La capa de objetos convierte en información a las agregaciones etiquetadas de tipos de datos primitivos procedentes de la capa de estructura, representándola como objetos con significado para la aplicación y a través de esta última para el usuario. Los formatos de objetos podrían ser abiertos o propietarios.
5. La capa de aplicación contiene software personalizado que manipula los objetos de la capa anterior y los presenta al usuario en una forma que los pueda entender.

Cada una de estas capas aportan elementos a las capas superiores para hacer que la información sea visualizada y entendida por el usuario, lográndose esto finalmente en la capa de aplicación. La transición por las diferentes capas expone los documentos a múltiples amenazas, por ejemplo en la capa física un dispositivo de almacenamiento puede tener un fallo quedando corrupta la secuencia de bits. Otro error puede ser que el formato de un documento sea incompatible con el sistema operativo o no sea reconocido por una aplicación, lo que provocaría pérdidas totales o parciales de la información. A esto se le une los problemas de seguridad, que es otra de las amenazas que afecta a los documentos digitales, como el borrado y cambio de información sin autorización. Si la información preservada está expuesta a estas amenazas, hace que la misma deje de ser auténtica para su utilización en cualquier esfera de la toma de decisiones.

### 1.3 Documento digital

Un documento se puede considerar según las definiciones del proyecto InterPares 2 como “una unidad indivisible de información constituida por un mensaje fijado a un medio (registrado) en una forma sintáctica estable”.

(InterPARES 2 Project, 2008)

“Un documento digital es información registrada en un formato que requiere un ordenador para procesarla y que, además, se ajusta a la definición de documento.”

(ISO, 2011)

“La mínima cantidad de información con significado, expresada en forma de dígitos generalmente binarios codificados de acuerdo a una convención preestablecida y fijados en un medio digital.”

(Internacional council on archives, 2010)

“En el caso de un documento digital, la información está codificada en bits, y para leer, visualizar o grabar la información se precisa de un dispositivo que transmita o grabe información codificada en bits. Al representarse digitalmente, los datos de entrada son convertidos en dígitos (0,1) inteligibles para la máquina y no para los sentidos humanos; y a la salida, otro dispositivo los convertirá en señales analógicas, inteligibles para los sentidos humanos. Un documento digital es, pues, aquel que contiene la información codificada en bits.”

(Lamarca Lapuente M., 2013)

De los conceptos analizados se puede concluir que un documento digital para poder ser visualizado requiere de un ordenador o dispositivo con capacidad de interpretar las cadenas de bits que lo componen. Esto lleva consigo la utilización del software y el hardware necesario para su procesamiento, interpretación y visualización.

Los documentos digitales se generan en los más diversos ambientes y esto trae consigo que se generen en diferentes formatos, lo que hace aún más complicada la tarea de preservarlos. A cada uno de los formatos se aplican técnicas de preservación diferentes y debido al volumen tan grande de información se tiene la necesidad de utilizar software que implementen estas técnicas de manera automática.

### **Cualidades de los documentos digitales**

1. Disponibilidad: La disponibilidad es la propiedad de que un objeto digital está disponible para su uso a largo plazo. Con el fin de garantizar la disponibilidad, el objeto digital debe ser ingerido en, y posteriormente mantenido por, un repositorio de preservación. Si bien pueden haber barreras físicas a esto, más a menudo se trata de una cuestión de la prioridad que los tomadores de decisiones se adhieren a su valor a largo plazo o los permisos otorgados por aquellos que controlan los derechos de propiedad intelectual relacionados con el objeto. (Vermaaten, S., Lavoie, B., & Caplan, P., 2012)
2. Persistencia: es la propiedad de continuidad de la existencia de las secuencias de bits que comprenden un objeto digital y que se puedan recuperar y procesar del medio en el que están almacenadas. A fin de que el objeto digital permanezca siendo útil en el tiempo, es esencial que las secuencias de bits no estén corrompidas en ninguna manera, y que puedan ser leídas en su totalidad de los medios físicos en los que se almacenan. (Vermaaten, S., Lavoie, B., & Caplan, P., 2012)
3. Interpretabilidad: es la propiedad por la cual un objeto digital es capaz de ser utilizado de una manera que conserva sus características significativas. Un entorno de hardware y software apropiado permite a los usuarios interactuar con (ver, escuchar, consultar) el objeto de una

manera que conserva las características del objeto original que se consideran importantes por los interesados. Para las obras experienciales esta propiedad adquiere mayor relevancia y está relacionada a la autenticidad, tal es el caso de las obras de arte. (Vermaaten, S., Lavoie, B., & Caplan, P., 2012)

4. **Comprensibilidad:** requiere asociar suficiente información suplementaria con el contenido digital archivado tal que el contenido pueda ser interpretado y entendido adecuadamente por los usuarios previstos. Aún si a un documento digital le son preservadas las demás propiedades, puede quedar inutilizable si su contenido no es comprensible. (Vermaaten, S., Lavoie, B., & Caplan, P., 2012)
5. **Integridad:** Cualidad de un documento estar completo e inalterado en todos sus aspectos esenciales. (Vermaaten, S., Lavoie, B., & Caplan, P., 2012)
6. **Identidad:** es la propiedad de poder ser referenciado. La identidad distingue un objeto de otros objetos en un grupo y permite que un objeto sea descubierto y recuperado. Una cantidad limitada de metadatos (por ejemplo, nombre, número de identificación único, fecha, número de versión, creador) es a menudo todo lo que se requiere para efectos de identificación y eliminación de ambigüedades. En el caso de los documentos de archivo, los metadatos asociados a la identidad son más complejos y completos ya que ayudan a documentar la autenticidad, cualidad estrechamente relacionada a la identidad. (Vermaaten, S., Lavoie, B., & Caplan, P., 2012)
7. **Autenticidad:** La característica de un documento de archivo que es lo que pretende ser y que está libre de manipulación o corrupción (InterPARES 2 Project, 2008)

Cuando se preservan los documentos digitales es necesario mantener sus cualidades esenciales de autenticidad e integridad, para poder ser utilizados como fuentes confiables de información.

La autenticidad se compone por la identidad y la integridad, si alguna de estas fallas, el documento no puede ser considerado auténtico y como consecuencia dejaría de ser una fuente segura de información.

### 1.4 Formatos

Los formatos de los documentos digitales tienen una gran importancia, ya que es la manera en la que se estructuran internamente. Como los formatos son cada vez más cambiantes existen un conjunto de estrategias para la preservación digital encaminadas directamente a tratar este fenómeno.

“El formato de fichero o formato digital es la organización de los datos dentro de los objetos digitales, generalmente diseñado para facilitar el almacenamiento, recuperación, procesamiento, presentación

y/o transmisión de los datos por el software. Los formatos responden a una forma particular en que la información digital es representada por las aplicaciones informáticas.”

(InterPARES 2 Project, 2008)

“El formato de un archivo es una forma particular de codificar información para ser almacenada. Existen diferentes tipos de formatos para diferentes tipos de información, por esto existen formatos de gráficos, formatos de audio, formatos de animación, formato de documentos, etc.”

(Alegsa, 2010)

Se debe tener en cuenta que los formatos se agrupan en: formatos de audio, formatos de video, formatos de imagen, formatos de texto, formatos de uso exclusivo, formatos del sistema y formatos de comprimidos, entre otros. Por lo general el formato solo es visible para el usuario mediante la extensión del archivo. Los tipos de formatos hoy exceden los 150, por lo cual es un reto para los sistemas de preservación identificarlos a todos y de esta manera aplicar estrategias como la migración y la emulación.

Por lo antes planteado es de vital importancia realizar una adecuada selección de los formatos de fichero, considerando algunos factores que afectan la viabilidad y costo de preservar la información con vista a los futuros cambios en el ambiente tecnológico (Library of Congress., 2015):

- **Divulgación:** se refiere al grado en que existen las especificaciones completas y herramientas para validar la integridad técnica y están accesibles a aquellos que producen y preservan los contenidos digitales.
- **Adopción:** se refiere al grado en que el formato ya es utilizado por los creadores primarios, divulgadores, o usuarios de los recursos de información.
- **Transparencia:** se refiere al grado en el que la representación digital está abierta al análisis directo con herramientas básicas, incluyendo la legibilidad humana utilizando solamente un editor de texto.
- **Auto-documentación:** los objetos digitales que son auto-documentados es probable que sean más fáciles de preservar en el largo plazo y menos vulnerables a las catástrofes que aquellos que se almacenan por separado de los metadatos necesarios para entender su contexto.
- **Dependencias externas:** se refiere al grado en el que un formato depende del hardware, el sistema operativo o software en particular para la representación o el uso.
- **Impacto de patentes:** las patentes relacionadas a un formato digital pueden inhibir la capacidad de las instituciones de archivo para mantener el contenido en ese formato.
- **Mecanismos de protección técnica:** la información digital a preservar a largo plazo no debe estar protegida por mecanismos técnicos como la encriptación, implementados en una forma que

impiden a los preservadores tomar las medidas apropiadas para preservar el contenido digital y hacerlo accesible para las futuras generaciones.

Cada uno de estos factores debe ser tenido en cuenta cuando se conforman las políticas de preservación digital en cada institución. Los formatos son fundamentales en las etapas del ciclo de vida de la información, durante el ingreso, el almacenamiento, la planificación de la preservación y el acceso, según lo planteado en el modelo OAIS, en correspondencia con las características del entorno tecnológico de los productores de la información digital, los preservadores y los usuarios finales.

### 1.5 Modelo OAIS

El modelo OAIS proporciona un marco para la comprensión y mayor conocimiento de los conceptos archivísticos necesarios para la preservación y el acceso de la información digital a largo plazo (CCSDS, 2012). Además, OAIS se ocupa de una amplia gama de funciones de preservación digital, incluyendo el ingreso, el almacenamiento, el acceso y la planificación de la preservación digital. También tiene en cuenta la migración de la información digital a los nuevos medios y formas, los modelos de datos utilizados para representar la información, el papel del software en la preservación y el intercambio de información digital entre archivos. Por otra parte, el modelo OAIS cuenta con características que lo hacen aplicable a cualquier institución que tenga entre sus responsabilidades la preservación a largo plazo de la información digital, y del mismo se derivan un conjunto de normas y estándares complementarios. El modelo no es una guía desde el punto de vista técnico del desarrollo de software, sino un marco conceptual que permite identificar las funcionalidades básicas que deben tener los sistemas para lograr la preservación digital, permitiendo además que el modelo no pierda generalidad, y pueda ser aplicado a una amplia gama de sistemas informáticos, sin importar la arquitectura y tecnologías que estos usen (Carrazana, 2015).

### 1.6 Vulnerabilidades y amenazas a la preservación digital

Cuando se habla de las amenazas a las que se exponen los documentos digitales es necesario analizar sus categorías, puesto que se debe tener en cuenta la manera en la que son enfrentadas por los diferentes sistemas de preservación. Autores como (Conway, 1996) caracterizan las actividades de preservación digital como procesos de gestión de riesgos, ya que el costo de la preservación es juzgado teniendo como base el costo del fracaso de preservar los documentos digitales, en términos de los usuarios que necesitan la información. En los últimos años, varios autores han definido los tipos de amenazas a la preservación de los contenidos digitales, entre ellos la de (Vermaaten, S., Lavoie, B., & Caplan, P., 2012) titulada *“Identifying Threats to Successful Digital Preservation: the SPOT Model for*

*Risk Assessment*”, donde se pueden encontrar los principales modelos de amenazas a las que se enfrenta la preservación digital.

En estos modelos se describen las amenazas teniendo en cuenta categorías generales para riesgos técnicos, legales y económicos; los mismos pueden servir como herramientas para realizar evaluaciones de riesgos. En el modelo descrito por (Barateiro, J., Antunes, G., & Borbinha, J., 2010) se define una clasificación de las vulnerabilidades y las amenazas a la preservación digital basada en la terminología usada en la gestión de riesgos; las vulnerabilidades son definidas como puntos de fallo potenciales y las amenazas como eventos que afectan el comportamiento normal del ambiente de preservación.

<b>Vulnerabilidades</b>	Datos	<ul style="list-style-type: none"> <li>• Fallo de los medios de almacenamiento</li> <li>• Obsolescencia de los medios de almacenamiento</li> </ul>
	Infraestructura	<ul style="list-style-type: none"> <li>• Fallas de hardware</li> <li>• Obsolescencia del hardware</li> <li>• Fallos de comunicación</li> <li>• Fallos en el servicio de red</li> </ul>
	Procesos	<ul style="list-style-type: none"> <li>• Fallos de software</li> <li>• Obsolescencia del software</li> </ul>
<b>Amenazas</b>	Desastres	<ul style="list-style-type: none"> <li>• Desastres naturales</li> <li>• Errores operacionales humanos</li> </ul>
	Ataques	<ul style="list-style-type: none"> <li>• Ataques internos</li> <li>• Ataques externos</li> </ul>
	Gestión	<ul style="list-style-type: none"> <li>• Fracasos económicos</li> <li>• Fracasos organizacionales</li> </ul>
	Legislación	<ul style="list-style-type: none"> <li>• Cambios legislativos</li> <li>• Requisitos legales</li> </ul>

**Tabla 1: Categorías de vulnerabilidades y amenazas a la preservación digital. Adaptada de** (Barateiro, J., Antunes, G., & Borbinha, J., 2010).

A continuación, se describen las vulnerabilidades y amenazas identificadas anteriormente (Rosenthal D., Robertson T., Lipkis T., Reich V., & Morabito S, 2005):

- Fallo de los medios de almacenamiento

Debe esperarse que los medios de almacenamiento se degraden con el tiempo, manifestándose fallos de discos, causando errores irreversibles en los bits y pérdida de grandes volúmenes de datos. Por lo general, los soportes utilizados para almacenar los objetos digitales son inestables y se deterioran al cabo de pocos años o décadas.

- Obsolescencia de los medios de almacenamiento y el hardware

Todos los medios y componentes de hardware eventualmente fallarán. Se convertirán en obsoletos en el sentido de no ser capaces de comunicarse con otros componentes del sistema o ser reemplazados cuando fallen. La utilización de los objetos digitales depende de medios de acceso que funcionan de manera específica, tratándose a menudo de combinaciones complejas de equipos y programas informáticos, que suelen dejar de estar vigentes en pocos años y son reemplazados por otros medios de funcionamiento distinto.

- Fallo de hardware

Debe esperarse que los componentes de hardware sufran fallas recuperables transitorias, como la pérdida de energía, y también fallas irreversibles catastróficas, tales como fuentes de alimentación quemadas.

- Obsolescencia de software

Se produce cuando un componente de software se vuelve obsoleto e incapaz de funcionar. Esto a menudo se manifiesta como obsolescencia de formato, aunque los bits en los que fueron codificados algunos datos permanecen siendo accesibles, la información ya no puede ser descodificada desde el formato de almacenamiento hacia una forma legible.

- Fallo de software

Los componentes de software son susceptibles de sufrir errores que representan un riesgo para los datos almacenados. Los procesos de negocio que para su realización se apoyan en servicios de software pueden verse afectados.

- Fallas de comunicación

Los sistemas no pueden asumir que las transferencias de red que utilizan para adquirir y diseminar contenido van a tener éxito o fracasar dentro de un periodo de tiempo especificado, o que en realidad entregarán el contenido inalterado.

- Fallas en el servicio de red

Los sistemas deben prever que los servicios externos de red que utilizan nombres de dominio y URL persistentes, van a sufrir fallos transitorios y no recuperables, tanto de los servicios de red como de entradas individuales en ellos.

- Desastres naturales



Los desastres naturales, tales como inundaciones, el fuego y terremotos deben ser anticipados ya que pueden causar daños a la información almacenada. Otros tipos de amenazas, tales como fallos en los medios de almacenamiento, de hardware y de infraestructura, típicamente se manifiestan.

- Errores operacionales humanos

Las acciones de los operadores pueden conducir a pérdidas de información, tanto recuperables como no recuperables. Esto no solo afecta a la aplicación informática para la preservación digital que se esté utilizando, sino también al sistema operativo en el que se está ejecutando, las otras aplicaciones que comparten el mismo entorno, el hardware subyacente y la red a través de la cual se comunican.

- Ataques externos

Las bibliotecas y archivos que utilizan papel están sujetos a ataques maliciosos; no hay ninguna razón para esperar que sus equivalentes digitales estén exentos. Peor aún, los sistemas conectados a las redes públicas son vulnerables a los virus y gusanos. Los sistemas de preservación digital deben defenderse contra los ataques, o estar completamente aislados de las redes externas.

- Ataques internos

En muchas de las violaciones a sistemas informáticos están implicadas personas con privilegios al mismo que le garantizan el acceso autorizado. Incluso si un sistema de preservación digital está completamente aislado de las redes externas, se debe anticipar el abuso de personas con acceso autorizado al mismo.

- Fracasos económicos

La información en forma digital es mucho más vulnerable a las interrupciones en el suministro de dinero que la información en papel. Hay costos actuales de energía, refrigeración, ancho de banda, la administración del sistema, registro de dominios, entre otros. Se debe esperar que los presupuestos para la preservación digital varíen en el tiempo y que incluso lleguen a agotarse.

- Fracasos organizacionales

Los fracasos organizacionales son la consecuencia de decisiones equivocadas que producen varias amenazas al ambiente de preservación. Desde el punto de vista del sistema de preservación digital, este debe incluir no solo la tecnología, sino la organización en la que está integrado. Estas organizaciones pueden morir, quizás a través de la quiebra, o sus misiones pueden cambiar. Esto puede privar a la tecnología de preservación digital de la ayuda que necesita para sobrevivir. La planificación del sistema debe prever la posibilidad de que el activo representado por el contenido preservado sea transferido a una organización sucesora, o de lo contrario ser eliminado adecuadamente.

- Legislación

Amenazas legislativas ocurren cuando los procesos de preservación digital o datos preservados violan una legislación nueva o revisada.

Estas amenazas atentan directamente contra las cualidades de los documentos para su preservación a largo plazo. La obsolescencia tecnológica puede impedir que la información de un documento sea leída, representada o comprendida, debido a incompatibilidades. De igual manera errores de funcionamiento del software pueden ocasionar pérdidas o corrupción de la información, afectando la integridad y autenticidad de los mismos.

Fallos de los medios de almacenamiento (discos duro, memorias flash, CD-ROM, etc.) traen consigo que la información no sea accesible, que ocurran pérdidas o los ficheros se corrompan. De igual manera ocurre con los problemas en el servicio de red que impiden que la información esté accesible mediante los sistemas informáticos para su búsqueda.

Los ataques internos y externos provocan diversos daños, uno muy común es la modificación o borrado de los documentos digitales, por lo cual pierden su autenticidad. Otra consecuencia es que la entidad responsable de la preservación digital dejaría de ser, al menos temporalmente, confiable para la custodia a largo plazo, hasta tanto las vulnerabilidades en su sistema de seguridad sean eliminadas y esto último corroborado por auditorías o certificaciones de un tercero confiable.

### **1.7 Estrategias de preservación digital**

Producto a las amenazas y vulnerabilidades antes mencionadas a las que se enfrentan la información en el entorno digital, se han propuesto un conjunto de estrategias, a continuación, se describen algunas de las más importantes.

- Migración

La migración es la transferencia periódica de materiales digitales de una configuración de software/hardware a otra, o de una generación de tecnología informática a una generación posterior. El propósito de la migración es el de preservar la integridad de los objetos digitales y conservar para los clientes la capacidad de recuperarlos, visualizarlos, y usarlos de cara al constante cambio de la tecnología. Sin embargo, la migración implica la pérdida de información en alguna medida, por lo que debe verificarse que si algún componente del documento se pierde, esto no afecte la capacidad del mismo para verificar su autenticidad en el tiempo (Duranti, 2001). Es más adecuada para objetos como imágenes, bases de datos o documentos de texto. Aún, cuando la información deja de estar en el entorno en que fue creada, perdiendo de esta manera cualidades de dicho entorno, se garantiza que la información esté accesible.

- Emulación

La emulación es la simulación en sistemas más recientes de las condiciones de ejecución de hardware y/o softwares originales en las cuales los objetos de información fueron concebidos (Barateiro, J.,

Antunes, G., & Borbinha, J., 2010). Este proceso es realizado por un programa llamado emulador, el cual se ejecuta en una computadora y virtualmente recrea una computadora diferente. (Hoeven, J., & Wijngaarden, H, 2005). Esta estrategia es comúnmente vista en juegos y algunas aplicaciones, donde la máquina recrea el ambiente tecnológico en que fueron creados.

- Diversidad

Las fallas en los sistemas de preservación digital están lejos de ser independientes. La diversificación de las propiedades de los componentes puede limitar el número de fallos simultáneos en el sistema. Algunas propiedades importantes que pueden ser diversificadas son (Barateiro, J., Antunes, G., & Borbinha, J., 2010):

- Ubicación física: diferentes ubicaciones geográficas pueden limitar el número de fallos simultáneos en el caso de desastres naturales o ataques a los componentes del sistema.
- Software: la diversificación de los sistemas operativos y otro software dispersan las vulnerabilidades a los gusanos o virus y también evita la dependencia del proveedor.
- Hardware: diversificar el hardware permite la reducción de la probabilidad de destrucción de componentes relacionados en condiciones específicas y también evita la dependencia de proveedores.
- Administración: la administración independiente evita que una sola persona comprometa todo el sistema de preservación digital en caso de un ataque interno o error humano.
- Almacenamiento: el uso de medios heterogéneos para el almacenamiento minimiza el impacto de los fallos debidos a defectos tecnológicos.
- Financiación: la diversificación de las fuentes de financiación de los sistemas de preservación digital evita trastornos económicos.

La diversidad permite que un sistema tenga una alta resistencia a fallos, por ejemplo, si se utilizan distintos sistemas operativos, podría ser atacado por un virus que afecte a Ubuntu, pero no tendría la misma repercusión si solo se ven afectados aquellos que tienen Ubuntu a que se afecten todos, porque utilizan Ubuntu.

- Inercia

La preservación digital es casi única entre las aplicaciones informáticas en que la velocidad no es ni un objetivo ni siquiera una ventaja. La falta de necesidad de velocidad se puede aprovechar para reducir el costo del hardware, energía y refrigeración. Pero la razón más importante para la inercia es que un sistema que funciona rápido tenderá a fracasar rápido, especialmente bajo ataque (Rosenthal D., Robertson T., Lipkis T., Reich V., & Morabito S, 2005). Esta estrategia es muy útil, por ejemplo, si se aplica en el proceso de eliminar la información, este será lento y si se sufre un ataque al sistema puede reducirse significativamente las pérdidas de información.

- Metadatos

Los metadatos permiten documentar de una manera estructurada la relación que existe entre el documento y sus diversos contextos, proporcionando evidencia del acto por el que el documento fue creado y en el que participa. Formalmente pueden definirse como “la documentación del contenido, contexto y estructura de un documento, que describe el proceso en que el documento participa, proporcionando de esta manera una pista de auditoría y conservación, identificando los agentes y eventos asociados con el documento y, en algunos casos, la descripción archivística” (Gilliland, 2005). Los metadatos tienen una gran importancia ya que son necesarios para la aplicación de estrategias como la migración y la emulación. También son claves al aportar la información necesaria para describir los requerimientos técnicos asociados al acceso, representación y uso de los documentos digitales.

- Auditoría

La auditoría apoya la detección de defectos latentes, permitiendo que el sistema se recupere más rápido y reduciendo la posibilidad de pérdidas. Por ejemplo, las anomalías que provocan la pérdida de datos solo se pueden detectar cuando se accede a los datos (Barateiro, J., Antunes, G., & Borbinha, J., 2010).

Esta estrategia está muy relacionada al chequeo de la integridad y autenticidad de los documentos digitales, teniendo técnicas de criptografía como la suma de chequeo (*o checksum en inglés*), el sellado de tiempo y la firma digital entre las más utilizadas.

Con la auditoría se puede verificar la integridad y autenticidad de los documentos digitales, detectándose y corrigiéndose las vulnerabilidades y amenazas si se utiliza esta estrategia de forma periódica, exceptuando los desastres naturales que no se detectan con la misma facilidad que el resto de las amenazas.

### 1.8 Descripción general de los sistemas de preservación

Los sistemas de preservación son herramientas informáticas diseñadas para garantizar que la información digital almacenada, sea cual sea el formato, programa, máquina o sistema utilizado para su creación, pueda permanecer y seguir utilizándose en el futuro pese a los rápidos cambios tecnológicos u otras causas que puedan alterar la información que contienen. Las contribuciones al campo de la preservación digital han sido realizadas principalmente desde áreas como las bibliotecas digitales, archivos digitales y repositorios digitales, teniendo como referencia más utilizada para la construcción de los sistemas de preservación digital al modelo *Open Archival Information System*, que ha dado lugar a la norma ISO 14721:2003 (Quisbert, 2008). Entre los sistemas de preservación digital a largo plazo se pueden encontrar los siguientes:

- Archivemática

Archivemática es un sistema de preservación digital gratuito y de código abierto desarrollado por Artefactual, diseñado para mantener el acceso a largo plazo a los documentos digitales. Está basado en web y ofrece un marco en el que los usuarios pueden procesar, preservar y acceder a los objetos digitales. Se estructura en torno al modelo de referencia OAIS. Cuenta con una extensa wiki para documentar sus características, funcionalidades y desarrollo, incluyendo tutoriales en video *screencast* y foros de usuarios. Por lo cual presenta un sistema de apoyo relativamente sólido para los usuarios que se están iniciando, o para la solución de problemas más avanzados. Todo el código de Archivemática es liberado bajo una Licencia Pública General de Affero (AGPL 3.0) que le da la libertad de estudiar, modificar, mejorar y distribuirlo. Es un sistema flexible, es decir los usuarios pueden configurar la mayoría de las opciones desde el ingreso hasta el almacenamiento de los archivos. (Archivemática, 2015)

- **LOCKSS (*Lots Of Copies Keep Stuff Safe*)**

Herramienta diseñada para las bibliotecas con el objetivo de asegurar la continuidad en el acceso de la comunidad a las revistas científicas publicadas en la web. Permite a las bibliotecas tomar la custodia del material al que se suscriban, de la misma manera que lo hace para el papel, y preservarlo. LOCKSS permite a las bibliotecas recoger contenido a medida que se publica y cooperar en una red par a par para detectar y reparar documentos dañados o ausentes. (Rosenthal D., Robertson T., Lipkis T., Reich V., & Morabito S, 2005). LOCKSS conserva los formatos web publicados (animaciones, bases de datos, imágenes, imágenes fijas, software, sonido, texto en movimiento) y géneros (revistas, libros, blogs, sitios web, archivos escaneados, audio, videos).

- **Kopal**

El sistema es un producto de la Biblioteca Kopal, implementado en Java y utiliza Kolibri como sistema operativo. Desarrollado para la interacción con el sistema DIAS de IBM dentro del proyecto Koppel. Este sistema trata con el problema de la obsolescencia de formatos de archivo a través de la migración de los formatos a lo largo de su arquitectura. Dentro de este sistema pueden ser manejadas varias versiones migradas de objetos y metadatos adicionales pueden describir cada proceso de migración. (Seadle, 2013)

- **Portico**

Es un servicio de preservación digital de revistas electrónicas, libros y otros contenidos, de ITHAKA, una organización sin fines de lucro dedicada a ayudar a la comunidad académica a utilizar las tecnologías digitales para preservar el registro académico y para avanzar en la investigación y la docencia de forma sostenible. Pórtico entiende la preservación digital como la serie de políticas y las actividades de gestión necesarias para garantizar la usabilidad, la autenticidad, la capacidad de descubrimiento y la disponibilidad de los contenidos en largo plazo. Sirve como un archivo permanente

de los contenidos de más de 117 editores, en nombre de más de 2.000 sociedades científicas y asociaciones (Morrissey, S. M., Meyer, J., & Bhattarai, S, 2010).

- **Preservica**

Preservica es un sistema de preservación digital estructurado en torno al modelo de referencia OAIS. Permite realizar controles de antecedentes de salud para proteger la integridad del contenido evitando así la pérdida y la corrupción. Este sistema puede ser instalado en el sistema operativo Windows, implementa estrategias como migración, auditoría y diversidad. Preservica está completamente probado, apoyado y mantenido por un equipo de desarrolladores de software profesionales y expertos en preservación digital. (Preservica, 2015)

### 1.9 Selección del Sistema de Preservación

Para realizar la selección del sistema se realizó una tabla de comparación donde se tuvieron en cuenta los siguientes aspectos:

- **Licencia**

Contrato entre el desarrollador de un software sometido a propiedad intelectual y a derechos de autor y el usuario, en el cual se definen con precisión los derechos y deberes de ambas partes. Es el desarrollador, o aquel a quien este haya cedido los derechos de explotación, quien elige la licencia según la cual distribuye el software.

(Gómez Labrador, septiembre de 2005)

- **Sistema operativo**

Conjunto de programas encargado de controlar los recursos del ordenador.

(Prieto A., 2002)

- **Lenguaje de programación**

Un lenguaje de programación es un convenio entre personas que puede definirse así:

Conjunto de reglas o normas que permiten asociar a cada programa correcto un cálculo que será llevado a cabo por un ordenador (sin ambigüedades). Por tanto, un lenguaje de programación es un convenio o acuerdo acerca de cómo se debe de interpretar el significado de los programas de dicho lenguaje.

(Ureña Almagro, 2011-2012)

- **Compatible con Dspace**

Producto a que el Repositorio institucional es una adaptación de Dspace, es necesario que el sistema de preservación que se seleccione sea compatible con el mismo. Esto facilitará la migración de datos.

- **Estrategias que implementan los sistemas**

## Capítulo 1: Aspectos teóricos de la preservación digital

Las estrategias de preservación de la información digital se utilizan para mitigar o enfrentar las posibles amenazas o vulnerabilidades que pueden afectar a los documentos digitales. En la siguiente tabla se muestra la comparación entre los sistemas teniendo en cuenta los criterios antes mencionados.

Sistema de preservación	Licencia	S.O	L.I	Compatible con Dspace	Estrategias que Implementan				
					M	E	A	D	I
Archivematica	AGPL 3.0	Ubuntu 12.04-14.02	PYTHON	Sí	X		X	X	
LOCKSS	Open Source	OpenBSD (2011)	JAVA	No	X		X	X	
Kopal	Open Source	Kolibri	JAVA	No	X				
Portico	LGPL	Windows	C++ y JAVA	Sí	X				
Preservica	Open Source	Windows	JAVA	Sí	X		X	X	

**Tabla 2: Comparación de los sistemas de preservación.**

### **Leyenda:**

- S.O** ---- Sistema operativo
- L.I** ---- Lenguaje de implementación
- M** ---- Migración
- E** ---- Emulación
- A** ---- Auditoría
- D** ---- Diversidad
- I** ---- Inercia

Como se puede apreciar en la tabla anterior se muestran varios sistemas de preservación de los más usados en el mundo. Cada uno de ellos resuelve el problema de la pérdida de información causado por las diferentes amenazas a las que puede estar sometido un documento digital, para ello hacen uso de algunas de las estrategias fundamentales de preservación. La selección del sistema a implantar se realizó teniendo en cuenta aspectos como sistema operativo, lenguaje de implementación, licencia, estrategias que implementan y su compatibilidad con Dspace.

Los sistemas Portico y Preservica son desarrollados utilizando software propietario. El uso de este tipo de software trae un conjunto de inconvenientes, ejemplo de esto pudiera ser que se terminen los fondos para pagar el uso de una licencia o la actualización de un software lo cual, si ocurre, pondría en riesgo a los documentos que se intentan preservar. A lo antes planteado, se debe añadir que es política de la UCI la utilización del software libre, por lo que se descartan los sistemas antes mencionados. De igual manera no se toma como solución el sistema Kopal, dado que se encuentra en una versión beta, la cual es una versión inestable donde pueden ocurrir cambios en el software, producto a correcciones de errores en el sistema, por lo cual no garantiza la preservación de la información.

En cuanto a LOCKSS y Archivematica, los mismos implementan las estrategias fundamentales para garantizar la disponibilidad y autenticidad de la información (Migración, Diversidad y Auditoría), además que son desarrollados con licencia y sistemas de libre distribución. Un elemento a tener en cuenta para realizar la selección es que el Repositorio Institucional con el que se gestiona la información científico-técnica de la Biblioteca de la UCI, es una adaptación de Dspace, por ello resulta necesario utilizar un sistema que se integre al mismo, lo cual no es garantizado por LOCKSS y sí por Archivematica.

Luego de analizados los aspectos antes mencionados, se concluye que Archivematica es el sistema que mejor se ajusta a las necesidades de preservación digital de la Biblioteca de la UCI. El mismo garantiza la compatibilidad a nivel de datos con el Repositorio Institucional, lo cual facilita la migración de los documentos digitales. Es de licencia libre, lo que implica un ahorro desde el punto de vista económico y que por tanto las tareas de preservación no estén en riesgo por cuestiones asociadas a la ausencia de recursos financieros. Producto a que la mayoría de los documentos están en los formatos pdf, doc y ppt, Archivematica asegura su disponibilidad, ya que implementa la migración, estrategia fundamental para preservar este tipo de formatos.

Archivematica cuenta además con una comunidad activa de usuarios y desarrolladores que lo mejoran periódicamente, lo cual da garantías de su continuidad en el futuro, evitando que este quede obsoleto.

### **1.10 Lenguaje, herramientas y tecnologías a utilizar.**

Al seleccionar Archivematica se hace necesario la utilización de un conjunto de herramientas utilizadas por este software, Python como lenguaje de implementación en su versión 2.7.4, Django 1.5.4 como framework, como servidor de bases de datos MySQL Server 5.5, como servidor de aplicaciones web Apache 2, además de utilizar los servidores Nginx y Gearman.



### 1.11 Metodología a utilizar

ITIL (*Information Technology Infrastructure Library*) Biblioteca de Infraestructura de Tecnologías de Información, es una metodología de desarrollo y gestión de servicios para proyectos y empresas, destinada a facilitar la entrega de servicios de tecnologías de la información (TI). Esta metodología fue desarrollada a petición del Gobierno del Reino Unido a finales de los 80 y recoge las mejores prácticas en la gestión de los Sistemas de Información. Es un conjunto de documentos donde se describen los procesos requeridos para la gestión eficiente y efectiva de los Servicios de Tecnologías de Información dentro de una organización. Son un conjunto de mejores prácticas y estándares en procesos para hacer más eficiente el diseño y administración de las infraestructuras de datos dentro de la organización. Es un “marco de trabajo” (framework) para la Administración de Procesos de TI. (BRAVO, y otros, 2006)

La metodología ITIL V3 está compuesta por cinco procesos (ITIL & IT SERVICE MANAGEMENT BOOKSHOP, 2014):

- Estrategia del Servicio (*Service Strategy*)
- Diseño del Servicio (*Service Design*)
- Transición del Servicio (*Service Transition*)
- Operación del Servicio (*Service Operation*)
- Perfeccionamiento Continuo del Servicio (*Continual Service Improvement*)

Dado que Archivematica es un sistema ya implementado y lo que se pretende es implantarlo en la Biblioteca de la UCI, de esta metodología solo se utilizará el proceso Transición del Servicio, por cubrir los procesos esenciales a realizar en la implantación.

El objetivo de este proceso es construir y desplegar servicios de tecnologías de información nuevos o modificados y asegurar que la gestión de los servicios es realizada de una manera coordinada. A continuación se describen los procesos que son parte de Transición del Servicio (Kempter, S., 2016):

- Gestión de Cambios: Controlar el ciclo de vida de todos los cambios. El objetivo primordial de la Gestión de Cambios es viabilizar los cambios beneficiosos con un mínimo de interrupciones en la prestación de servicios de TI.
- Evaluación de Cambios: Permite evaluar los cambios principales, tales como la introducción de un nuevo servicio o un cambio substancial a un servicio existente, antes de que se permita proceder a realizar los cambios en la próxima fase del ciclo de vida.
- Desarrollo de aplicaciones: Hacer disponible aplicaciones y sistemas los cuales provean la funcionalidad requerida para los servicios de tecnologías de la información. Este proceso incluye la personalización de productos de proveedores de software.

- Gestión de liberaciones y despliegue: Permite planificar y controlar el movimiento de liberaciones para pruebas y entornos reales. La meta principal de este proceso es asegurar que la integridad del entorno real está protegida y que los componentes correctos están liberados.
- Validación y Pruebas de Servicios: Asegura que las liberaciones desplegadas y los servicios resultantes cumplen las expectativas de los clientes, y verifica que las operaciones de TI son capaces de soportar los nuevos servicios.

Aunque la metodología seleccionada no describe procesos de desarrollo de software, se adapta correctamente a los requerimientos del objetivo de la presente investigación y puede ser complementada con metodologías de desarrollo de software existentes.

Es necesario tener en cuenta que al contar con un producto de software como Archivematica, completamente desarrollado y probado por una comunidad internacional especializada en temas de preservación digital, el mayor esfuerzo estará encaminado a adaptar el software a las necesidades de la biblioteca, para su despliegue y uso. Estas actividades son cubiertas correctamente por el proceso de Transición del servicio con cada uno de sus subprocesos.

Los procesos antes descritos serán aplicados como guía para realizar la implantación del sistema de preservación digital en la Biblioteca de la UCI.

Luego de culminado el capítulo y de realizar un estudio preliminar sobre los elementos básicos de preservación digital se concluyó que:

- El Repositorio institucional que está basado en Dspace posee un conjunto de limitaciones que no garantizan la perduración de la información a largo plazo, ni cualidades esenciales como autenticidad y disponibilidad.
- Es necesario implantar un sistema de preservación digital en la Biblioteca de la UCI para salvaguardar la información científico-técnica en formato digital que esta custodia.
- El sistema más adecuado para ser implantado en la Biblioteca de la UCI es Archivematica y como consecuencia se heredaron el lenguaje, las herramientas y las tecnologías de este.

## **Capítulo 2: Propuesta de solución**

La implantación es el proceso de instalar un nuevo software, que da como resultado la sustitución o mejoramiento de las actividades de forma automatizada. En la presente investigación se propone como solución implantar en la Biblioteca de la UCI el sistema Archivematica, el cual automatiza las tareas más importantes asociadas a la preservación digital a largo plazo.

La implantación<sup>1</sup> va más allá de una simple instalación de software, pues no solo se propone dejar en ejecución el sistema Archivematica, sino que afectará los procesos de la organización al incluirse nuevas tareas y recursos materiales y humanos para realizar la preservación digital.

Para realizar la implantación se han desarrollado las siguientes tareas: análisis de los principales sistemas de preservación digital existentes en el mundo, elección de uno de los sistemas analizados, instalación, configuración, internacionalización al español, migración de los documentos digitales, despliegue y capacitación. Las primeras dos tareas fueron expuestas en el capítulo uno de la investigación, mientras que las restantes serán presentadas en el actual capítulo.

Cada una de las tareas mencionadas anteriormente se ha realizado teniendo en cuenta los procesos asociados a Transición del servicio de la metodología ITIL, descritos en el capítulo 1, como marco para aplicarlas de una manera coordinada y coherente.

### **2.1 Evaluación de Cambios**

Para implantar exitosamente el sistema propuesto, se deben tener en cuenta aspectos esenciales del contexto que pueden influir en las acciones a realizar para garantizar la preservación digital. Los cambios esenciales están relacionados con la introducción de un nuevo servicio en la Biblioteca de la UCI, ya que la misma no tiene entre sus objetivos estratégicos la preservación digital. Por tanto, se introduce la preservación digital como un nuevo servicio en la biblioteca, el cual estará soportado en las funcionalidades que brinda Archivematica. Este nuevo servicio provocará, además, el surgimiento de nuevos procesos en la organización, así como la asignación de nuevos recursos para realizar las tareas necesarias para la preservación.

Otra perspectiva a considerarse en cuanto a los cambios más importantes a realizar, es la relacionada al uso del Repositorio Institucional que, si bien se usa hasta hoy como herramienta de gestión, debe

---

<sup>1</sup> Establecer y poner en ejecución nuevas doctrinas, instituciones, prácticas o costumbres. (Real Academia Española, 2016)

ahora pensarse también en su integración con el sistema de preservación digital a implantar, por lo cual una migración de los documentos hacia este último sistema es esencial.

El Repositorio Institucional contiene actualmente un total de 7894 documentos en los formatos pdf, doc y ppt, que comenzaron a almacenarse desde el año 2001 y esta cantidad aumenta con el transcurso de los años. Las características que tienen los formatos utilizados determinan las estrategias de preservación a utilizar. Dado que estos documentos no son recursos digitales interactivos como software, multimedia o juegos, la migración resulta ser la estrategia más adecuada en este caso. El hecho de que aún no se cuente con un gran volumen de información permite que se puedan utilizar máquinas de bajas prestaciones para instalar en la Biblioteca el sistema propuesto en esta investigación.

## **2.2 Desarrollo de aplicaciones**

Aunque Archivematica es el sistema de preservación seleccionado para realizar la implantación, aún quedan pendientes algunas tareas de personalización del mismo, necesarias para su adaptación a las necesidades de la Biblioteca de la UCI.

La personalización incluye básicamente la configuración de los elementos de la administración del sistema, internacionalización al español y la migración de los documentos digitales almacenados en Dspace. Todas estas actividades se han desarrollado en un entorno de desarrollo, para posteriormente desplegar el sistema en el entorno real en que será utilizado.

Para lograr una mejor comprensión de los elementos necesarios a tener en cuenta para realizar la personalización de Archivematica, se describen en los siguientes acápite su arquitectura y principales funcionalidades.

### **2.2.1 Arquitectura de Archivematica**

La estructura de Archivematica implementa un enfoque de micro-servicios, permitiendo que se realicen pequeñas tareas o servicios de forma automática y que se comuniquen los paquetes de información de OAIS entre ellos utilizando peticiones HTTP. Archivematica también está conformado por un *Dashboard* (Tablero de instrumentos web) que les permite a los usuarios procesar, supervisar y controlar los procesos de flujo de trabajo del sistema. La utilización de este tablero de instrumento proporciona una interfaz multi-usuario, que presenta un informe sobre el estado de los eventos del sistema y hace que sea más fácil de controlar y activar un micro-servicio específico.

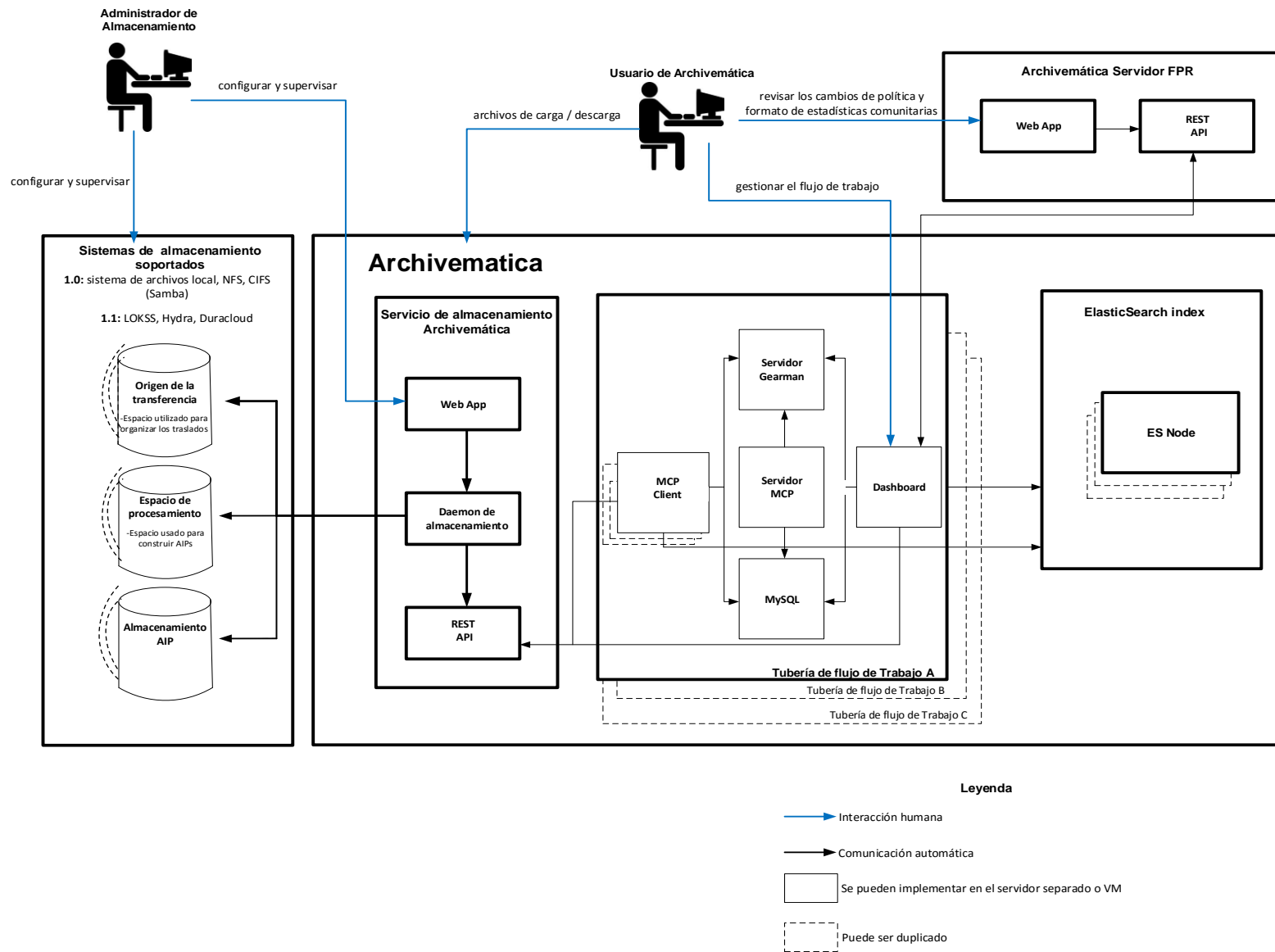


Figura 3: Diagrama de la arquitectura de Archivemática. Adaptado de (Artefactual, 2016).

### Servicio de almacenamiento

Archivematica permite que el soporte de almacenamiento este implementado en un servidor aparte o en una máquina virtual. El administrador se encarga de configurar y supervisar el espacio que se utiliza para organizar las transacciones, el espacio de procesamiento y el de almacenamiento.

Los paquetes PIAs contienen los metadatos METS.xml<sup>2</sup> y los documentos, esto permite no solo preservar los documentos, sino mantener registrados los elementos necesarios para su preservación. Los metadatos contenidos en los paquetes permiten comprobar la autenticidad de los documentos y aportan elementos necesarios para realizar estrategias como la migración, con la cual se garantiza la disponibilidad de la información digital.

El usuario de Archivematica no puede eliminar ningunos de los PIAs almacenados, él hace una solicitud y el administrador del sistema de almacenamiento es quien la revisa y decide si el paquete debe ser eliminado. Al ser eliminado un PIA se genera un registro que contiene: un número único de identificación, la razón del que realiza la solicitud y la del administrador, el usuario que realizó la solicitud y la fecha.

### Dashboard

De igual manera el sistema de preservación internamente permite que se gestione el servicio de almacenamiento, *ElasticSearch* y los flujos de trabajo de forma independiente.

- **El servicio de almacenamiento** es configurado y supervisado por el administrador mediante la web, esta se comunica con el *Daemon* de almacenamiento que a su vez utiliza para sus procesos el espacio disponible en el soporte de almacenamiento y emplea *REST API* para la transferencia de datos.
- **Todo el flujo de trabajo** puede realizarse de forma independiente, se utiliza MCP (*Mud Client Protocol*) que es un protocolo que está dirigido para ser fácil de usar debido a que se integra bien con un servidor existente. MCP especifica un mecanismo para el intercambio de mensajes entre el cliente y el servidor, pero no los detalles de esos mensajes, que se dejan al diseñador de la aplicación. Este protocolo es el núcleo del sistema Archivematica ya que controla los diversos micro-servicios en el sistema. Los clientes MCP le informan al servidor *Gearman* las tareas que pueden realizar y esperan a que el servidor les asigne una tarea.

---

<sup>2</sup> Metadata Encoding and Transmission Standard (METS).

- La configuración y el procesamiento de la información se llevan a cabo en la base de datos. Los usuarios monitorean y controlan el MCP a través del *Dashboard*. Este proporciona una interfaz multi-usuario basada en la web que informará sobre el estado de los eventos del sistema y hace que sea más fácil de controlar y activar el micro-servicio específico. Archivemática da seguimiento de la ubicación y existencia de un PIA dentro del **ElasticSearch** que se puede buscar utilizando el *Dashboard*.

**Servidor de políticas de formato**

El usuario de Archivemática es el responsable de revisar si hay nuevas actualizaciones de las políticas y formatos mediante la web. De encontrar cambios puede actualizar el servidor local desde el servidor de Registro Política de Formato o *Format Policy Registry* (FPR) que se encuentra en Internet, usando el botón Actualizar en el módulo Planificación de Preservación de Archivemática. Al aplicar correctamente las políticas de formato y mantenerlas actualizadas se garantiza que los documentos digitales se encuentren disponibles a largo plazo, debido a que estas políticas indican las acciones, herramientas y ajustes que se aplican a un objeto digital de un formato determinado. En la siguiente tabla se muestran los formatos que se preservan en Archivemática y las herramientas que utiliza.

<b>Tipos de Medios.</b>	<b>Formatos de archivos.</b>	<b>Formatos de preservación.</b>	<b>Formatos de acceso.</b>	<b>Herramientas de Normalización.</b>
Audio	AC3, AIFF, MP3, WAV, WMA	WAVE (LPCM)	MP3	FFmpeg
Email	PST	MBOX	MBOX	readpst
Email	Maildir**	Formato Original	MBOX	md2mb.py
Office Open XML	DOCX, PPTX, XLSX	Formato Original	Formato Original	Herramienta de búsqueda en curso
Texto plano	TXT	Formato Original	Formato Original	Ninguna

Formato de Documento Portable	PDF	PDF/A	Formato Original	Ghostscript
Los archivos de presentaciones	PPT	Formato Original	PDF	Herramienta de búsqueda en curso
Imagen	BMP, GIF, JPG, JP2*, PCT, PNG*, PSD, TIFF, TGA	TIFF descomprimido	JPEG	ImageMagick
Archivos RAW de cámara / formato de negativo digital**	3FR, ARW, CR2, CRW, DCR, DNG, ERF, KDC, MRW, NEF, ORF, PEF, RAF, RAW, X3F	Formato Original	JPEG	ImageMagick/UFRaw
Las hojas de cálculo	XLS	Formato Original	Formato Original	Ninguna
Las imágenes vectoriales	AI, EPS, SVG	SVG	PDF	Inkscape
Video	AVI, FLV, MOV, MPEG-1, MPEG-2, MPEG-4, SWF, WMV	FFV1/LPCM en MKV	MP4	FFmpeg
Archivos de procesamiento de textos	DOC, WPD, RTF	Formato Original	Formato Original	Herramienta de búsqueda en curso**

**Tabla 3: Formatos y herramientas utilizados para preservar en Archivematica.** (Archivematica, 2015)

(\*) JPEG y JPEG 2000 no se normalizan a un formato de conservación

(\*\*) En el desarrollo



De forma general Archivemática proporciona un conjunto integrado de herramientas libres y de código abierto que permite a los usuarios procesar objetos digitales desde el ingreso hasta el acceso de acuerdo con el modelo funcional ISO-OAIS.

### 2.2.2 Patrón de diseño de Archivemática

Archivemática utiliza un patrón de diseño de micro-servicios para proporcionar un conjunto integrado de herramientas de software que permite a los usuarios procesar objetos digitales desde el ingreso hasta el acceso de acuerdo con el modelo funcional ISO-OAIS. Archivemática utiliza METS, PREMIS, *Dublin Core* y otros estándares de mejores prácticas de metadatos. Los micro-servicios en Archivemática son tareas del sistema granulares que operan en una entidad conceptual que es equivalente a un paquete de información OAIS. Estos paquetes de información se procesan mediante una serie de micro-servicios, que son proporcionados por una combinación de secuencias de comandos Python y otras herramientas de software libre, que están dentro de los paquetes del sistema Archivemática. Cada micro-servicio termina en un estado de éxito o de error y el paquete de información se procesa en consecuencia por el siguiente micro-servicio. (Archivemática, 2015)

### 2.2.3 Funcionalidades de Archivemática

Archivemática se basa en el modelo OAIS (*Open Archives Information System*) analizado en el capítulo uno de la presente investigación, por eso cada uno de sus módulos coincide con los propuestos por este modelo.



Figura 4: Modelo OAIS. Adaptado de (CCSDS, 2012).

Módulo Transferencia

En Archivemática la transferencia es el proceso de transformar cualquier conjunto de objetos digitales o directorios en un PIE<sup>3</sup>. Este proceso puede incluir la evaluación, ordenación, descripción e identificación de los donantes restringidos, privados o contenidos confidenciales. La transferencia prepara a un objeto digital para su preservación en Archivemática. En la transferencia el usuario mueve los objetos digitales de directorios de fuentes accesibles a través del servicio de almacenamiento en Archivemática. Una vez cargado en el *Dashboard*, las transferencias se ejecutan a través de una serie de micro-servicios. Al final de la transferencia, el usuario crea un PIE de una transferencia, o elige para enviar la transferencia, donde se puede dividir en diferentes PIE, o combinadas en un SIP con otras transferencias.

La siguiente tabla muestra los micro-servicios más importantes del módulo transferencia y cómo permiten garantizar la autenticidad (identidad, integridad) y disponibilidad de los documentos a largo plazo.

<b>Micro-servicio</b>	<b>Descripción</b>	<b>Aporte a la preservación digital</b>
<p>Cambiar el nombre de transferencia con UUID<sup>4</sup>.</p> <p>Tareas más importantes:</p> <ul style="list-style-type: none"> <li>• Cambiar el nombre de transferencia con UUID.</li> </ul>	<p>Está directamente asociado con la transferencia de sus metadatos añadiendo el UUID de traslado al nombre del directorio de transferencia.</p>	<p>Este micro-servicio garantiza la identidad de los documentos ya que aporta un identificador único a la transferencia que la diferencia de las demás.</p>
<p>Asignar UUID de archivos y sumas de comprobación.</p> <p>Tareas más importantes:</p> <ul style="list-style-type: none"> <li>• Asignar UUID de archivos a los objetos.</li> <li>• Asignar sumas de comprobación y tamaño de los archivos a los objetos.</li> </ul>	<p>Asigna un identificador único universal y emplea el algoritmo SHA-256 en la suma de comprobación para cada archivo del directorio /objects/ y establece permisos de archivo para permitir la continuación de la tramitación.</p>	<p>En este micro-servicio se garantiza la identidad e integridad de los documentos. Cuando se le asigna un identificador único a cada documento se asegura que sea diferente y al asignarle las sumas de comprobación se verifica que el documento sea</p>

<sup>3</sup> Paquete de Información de Envío (PIE): es el paquete que es enviado al archivo por el productor. Su forma y contenido detallado son negociados entre el productor y la administración del archivo. Es usado en la construcción o actualización de uno o más paquete de información de archivo. (CCSDS, 2012)

<sup>4</sup> Unique Universal Identifier (UUID): Identificador universal único.

		íntegro y por consecuencia auténtico.
Verificar las sumas de comprobación de la transferencia.	Comprueba los archivos de suma de comprobación que se colocaron en la carpeta /metadata/ de la transferencia antes de mover la transferencia a Archivematica.	La sumas de comprobación garantiza la integridad de los documentos, en este micro-servicio se compara la suma de comprobación proporcionada por el productor con la calculada en el sistema, si es la misma el documento es transferido, de lo contrario se rechaza.
Elaboración de un documento METS.xml.	Genera un archivo METS básico con un fileSec y structMap para registrar la presencia de los objetos del directorio /objects/ y sus ubicaciones en los subdirectorios. Diseñado para capturar el orden original de la transferencia en el caso de que el usuario elija posteriormente eliminar, renombrar, mover archivos o deshacer la transferencia en múltiples PIEs. Una copia del archivo METS se añade automáticamente a cualquier PIE generado a partir de la transferencia.	Este micro-servicio garantiza la integridad y disponibilidad de los documentos digitales debido a que los archivos METS guardan los elementos necesarios para poder ser preservados.
Análisis de virus.	Utiliza ClamAV para escanear en busca de virus y otros programas maliciosos. Si se detecta un virus, la transferencia se coloca automáticamente en /sharedDirectoryStructure/failed/ y	Este micro-servicio evita que un documento con virus sea preservado y de esta manera garantiza la integridad y disponibilidad de los documentos.

	todo el procesamiento de la transferencia se detiene.	
Limpiar los nombres. Tareas más importantes: <ul style="list-style-type: none"> <li>Desinfectar archivo de objeto y nombres de directorio.</li> <li>Desinfectar nombre de Transferencia.</li> </ul>	Algunos sistemas de archivos no son compatibles con Unicode u otros caracteres especiales en los nombres de archivo. Este micro-servicio elimina caracteres prohibidos y los reemplaza con guiones, los nombres originales de archivo se conservan en los metadatos PREMIS.	Este micro-servicio garantiza que el documento no quede inaccesible por contener caracteres extraños en su nombre o en el de la transferencia. Por lo que garantiza la disponibilidad.
Identificar el formato del archivo.	Identifica los formatos de los objetos de la transferencia utilizando ya sea FIDO <sup>5</sup> o una extensión de archivo basado en la elección del usuario. Los tipos de formatos se gestionan en el Registro Política de Formato. Este micro-servicio puede ser omitido y hecho en lugar de ingreso.	La identificación de los formatos de los documentos garantiza su disponibilidad, dado que es uno de los factores fundamentales para poder realizar la migración de formatos.
Caracterizar y extraer los metadatos.	Caracteriza y valida los formatos y los objetos extraídos de metadatos utilizando el archivo de información del sistema de herramienta (FITS).	La validación de los formatos de los documentos garantiza su disponibilidad, dado que es uno de los factores fundamentales para poder realizar la migración de formatos.

**Tabla 4: Micro-servicios del módulo Transferencia.**

**Funcionalidades:**

1. Iniciar transferencia
2. Editar metadato
3. Visualizar metadato
4. Aprobar la transferencia

<sup>5</sup> Format Identification for Digital Objects (FIDO): Herramienta para la identificación de objetos digitales.

5. Rechazar la transferencia
6. Crear un PIE y continuar con el procesamiento
7. Enviar transferencia al retraso acumulado

Módulo Ingreso

Proporciona los servicios y funciones para aceptar los paquetes de información de envío desde los productores (o desde elementos internos bajo control de la Administración) y preparar el contenido para el almacenamiento y la gestión dentro del archivo. Las funciones del módulo Ingreso incluyen recibir y evaluar la calidad de los paquetes de información de envío, generar un paquete de información de archivo que se ajuste a los estándares de documentación y formato de datos del archivo, extraer su información descriptiva para incluirla en la base de datos y coordinar actualizaciones del Almacenamiento de archivos y la Gestión de datos. (CCSDS, 2012)

La siguiente tabla muestra los micro-servicios más importantes del módulo ingreso y cómo permiten garantizar la autenticidad (integridad, identidad) y disponibilidad de los documentos a largo plazo.

<b>Micro-servicio</b>	<b>Descripción</b>	<b>Aporte a la preservación digital</b>
Cambiar el nombre de directorio de PIE añadiendo un UUID. Tareas más importantes: <ul style="list-style-type: none"> <li>• Cambiar el nombre de directorio de PIE con PIE UUID.</li> </ul>	Directamente asocia el PIE con sus metadatos añadiendo el PIE UUID al nombre del directorio PIE y PIE comprueba si es del tipo de transferencia Maildir para determinar el flujo de trabajo.	Este micro-servicio garantiza la identidad de los documentos ya que aporta un identificador único al ingreso que lo diferencia de los demás.
Limpiar los nombres.	Algunos sistemas de archivos no son compatibles con Unicode u otros caracteres especiales en los nombres de archivo. Este micro-servicio elimina caracteres prohibidos y los reemplaza con guiones, los nombres originales de archivo se conservan en los metadatos PREMIS.	Este micro-servicio garantiza que el documento no quede inaccesible por contener caracteres extraños en su nombre o en el de la transferencia. Por lo que garantiza la disponibilidad.
Normalizar. Tareas más importantes:	Determina qué opciones están disponibles para la normalización del	Este micro-servicio permite la

<ul style="list-style-type: none"> <li>• Normalizar.</li> <li>• Normalizar para el acceso.</li> <li>• Normalizar para la preservación.</li> </ul>	<p>PIE y los presenta al usuario como opciones. Normaliza (para la preservación y el acceso) en base a la selección.</p>	<p>identificación y migración de los formatos, lo que garantiza la disponibilidad de los mismos.</p>
<p>Documentación de presentación de Proceso.</p> <p>Tareas más importantes:</p> <ul style="list-style-type: none"> <li>• Asignar UUID de archivos a la documentación de presentación.</li> <li>• Asignar sumas de comprobación y tamaño de los archivos de documentación de presentación.</li> <li>• Desinfectar los nombres de los archivos y directorios en la documentación de presentación.</li> <li>• Análisis de virus en la documentación de presentación.</li> <li>• Caracterizar y extraer los metadatos en la documentación de presentación.</li> </ul>	<p>Procesa toda la documentación de presentación incluido en el PIE y lo añade al directorio /objects/.</p>	<p>Este micro-servicio garantiza la identidad, la integridad y la disponibilidad de los documentos, por ejemplo, cuando se establecen nombres únicos se garantiza la identidad y al asignar las sumas de comprobación la integridad. Al eliminar los caracteres extraños de los nombres de los archivos y directorios se garantiza la disponibilidad. Analizar si los documentos contienen virus es una forma de garantizar que los mismos sean íntegros y que estén disponibles.</p>
<p>Procesos del directorio de metadatos.</p> <p>Tareas más importantes:</p>	<p>Procesos de metadatos.</p>	<p>Este micro-servicio garantiza la identidad, la integridad y la</p>

<ul style="list-style-type: none"> <li>• Asignar UUID de archivos a los metadatos.</li> <li>• Asignar sumas de comprobación y tamaño del archivo de metadatos.</li> <li>• Desinfectar los nombres de los archivos y directorios en los metadatos.</li> <li>• Buscar virus en los metadatos.</li> </ul>		<p>disponibilidad de los documentos, por ejemplo, cuando se establecen nombres únicos se garantiza la identidad y al asignar las sumas de comprobación la integridad. Al eliminar los caracteres extraños de los nombres de los archivos y directorios se garantiza la disponibilidad. Analizar si los documentos contienen virus es una forma de garantizar que los mismos sean íntegros y que estén disponibles.</p>
<p>Preparar PIA.</p> <p>Tareas más importantes:</p> <ul style="list-style-type: none"> <li>• Verificar las sumas de comprobación que se generan en el ingreso.</li> <li>• Elaboración de un documento METS.xml.</li> </ul>	<p>Crea un PIA en formato Bagit, el archivo puntero del PIA e índices del PIA sin pérdida de datos al comprimirlos.</p>	<p>Este micro-servicio garantiza la integridad y disponibilidad de los documentos, ya que los archivos METS guardan los elementos necesarios para poder preservar los documentos digitales. La suma de comprobación garantiza la integridad de los documentos.</p>

**Tabla 5: Micro-servicios del módulo Ingreso.**

### **Funcionalidades:**

8. Crear un PIE
9. Adicionar metadato
10. Editar metadato
11. Normalizar para la preservación y el acceso
12. Normalizar para la preservación
13. Rechazar PIE
14. No realizar normalización
15. Normalizar manualmente
16. Normalizar para acceso
17. Organizar PIE
18. Adicionar directorio

### Módulo Almacenamiento de archivos

Proporciona los servicios y funciones para el almacenamiento, mantenimiento y recuperación de los paquetes de información de archivo. Las funciones del módulo Almacenamiento de archivo incluyen recibir estos paquetes desde el ingreso y agregarlos al almacenamiento permanente, la gestión de la jerarquía de almacenamiento, el refrescamiento de los medios en los cuales los recursos del archivo son almacenados, la realización de chequeos de errores de rutina, proporcionando capacidades de recuperación ante desastres, y la entrega de los paquetes de información de archivo a la entidad funcional de Acceso para satisfacer los pedidos de los usuarios. (CCSDS, 2012)

### **Funcionalidades:**

19. Abrir PIA<sup>6</sup>
20. Descargar PIA
21. Eliminar PIA
22. Buscar un PIA almacenado
23. Eliminar campos de búsqueda
24. Visualizar todos los archivos

Para comprobar la integridad de los PIA en el almacenamiento, Artefactual ha desarrollado una aplicación llamada *Fixity*. (Artefactual, 2016)

### Módulo Administración

Provee los servicios y funciones para la operación de todo el sistema de archivo. Incluye entre sus funciones solicitar y negociar acuerdos de envío con los productores, auditar los envíos para garantizar que cumplen los estándares del archivo y el mantenimiento de la gestión de la configuración del

---

<sup>6</sup> Paquete de Información de Archivo (PIA): es el paquete de información que es preservado y debe conformarse a los estándares internos de preservación del archivo. (CCSDS, 2012)



hardware y software del sistema. También provee funciones de ingeniería de sistema para monitorear y mejorar las operaciones del archivo, así como llevar un inventario e informar sobre las migraciones y actualizaciones del contenido del archivo. (CCSDS, 2012)

### **Funcionalidades:**

25. Configurar para el procesamiento de una transferencia
26. Configurar las opciones de la interfaz
27. Cargar la url del servicio de almacenamiento
28. Salvar la url
29. Visualizar fallos de transferencia
30. Visualizar reportes de fallos
31. Eliminar transferencias fallidas
32. Visualizar localización de una transferencia
33. Visualizar localización de un PIA
34. Visualizar espacio para el procesamiento
35. Borrar contenido almacenado en los espacios de procesamiento
36. Adicionar PID en línea
37. Configurar REST PIA
38. Adicionar usuario al sistema
39. Editar usuario
40. Eliminar usuario del sistema

### Módulo Planificación de la Preservación

Provee los servicios y funciones para monitorear el entorno del archivo, brindando recomendaciones y planes de preservación para garantizar que la información almacenada permanezca accesible y entendible por la comunidad designada a largo plazo, aún si el ambiente tecnológico original queda obsoleto. Entre sus funciones se encuentran la evaluación del contenido del archivo, recomendar la migración de los recursos actuales de archivo, brindar periódicamente reportes de análisis de riesgos y monitorear los cambios sobre el ambiente tecnológico y los requerimientos de la comunidad designada. (CCSDS, 2012)

### **Funcionalidades:**

41. Crear nuevo formato
42. Visualizar formato
43. Editar formato
44. Actualizar formato
45. Buscar
46. Eliminar grupo de formato

47. Crear nuevo grupo de formato
48. Editar grupo de formato
49. Crear nueva herramienta
50. Visualizar nueva herramienta
51. Editar nueva herramienta
52. Crear nueva regla
53. Visualizar nueva regla
54. Reemplazar nueva regla
55. Deshabilitar nueva regla
56. Crear nuevo comando
57. Visualizar nuevo comando
58. Reemplazar nuevo comando
59. Deshabilitar nuevo comando

#### Módulo Acceso

Provee los servicios y funciones que apoyan a los consumidores en la determinación de la existencia, localización y disponibilidad de la información almacenada en el archivo. Entre sus funciones se incluyen la comunicación con el consumidor para recibir solicitudes, generar respuestas a través de paquetes de información de diseminación y aplicar controles para limitar el acceso a información especialmente protegida. (CCSDS, 2012)

#### **Funcionalidades:**

60. Eliminar registro del PID <sup>7</sup>

#### **2.2.4 Personalización de la interfaz**

Archivematica originalmente está en idioma inglés y no cuenta con ninguna versión o funcionalidad en español, por lo que entre los cambios realizados al sistema está su internacionalización, para un mejor entendimiento por parte de los usuarios del sistema. Se realizó la internacionalización accediendo al código fuente de la aplicación, donde se modificaron los ficheros html, algunos en python y otros en css; resultando ser más ciento veinte archivos en total. Los componentes del sistema que fueron traducidos forman parte el *DashBoard*, que se compone por los módulos transferencia, ingreso, planificación de la preservación, administración, acceso y el servicio de almacenamiento.

---

<sup>7</sup> Paquete de información de diseminación (DIP): es un paquete de información, que se deriva de una o más paquetes de información de archivo, enviados al Consumidor en respuesta a sus consultas (CCSDS, 2012).



Figura 5: Dashboard en español.

### 2.2.5 Migración de datos

Una etapa indispensable para realizar la implantación del sistema propuesto es la migración de los documentos digitales existentes en el Repositorio Institucional. Se consideraron varias alternativas para realizar el intercambio de datos entre los dos sistemas, entre ellas el acceso a los datos del repositorio través de OAI-PMH y OAI-ORE, pero las mismas fueron descartadas. En el primer caso, los metadatos obtenidos fueron insuficientes ya que las peticiones realizadas solo devolvían información de ciento uno registros asociados a documentos digitales. En el segundo caso la versión actual de Dspace no está soportando correctamente OAI-ORE, lo cual será corregido en actualizaciones futuras al sistema. La situación anterior fue además corroborada con un especialista del proyecto de desarrollo de Dspace de la UCI.

Finalmente, la integración entre los dos sistemas se realizó a nivel de datos teniéndose dos tareas fundamentales: la exportación de los documentos en Dspace y su posterior importación en Archivematica.

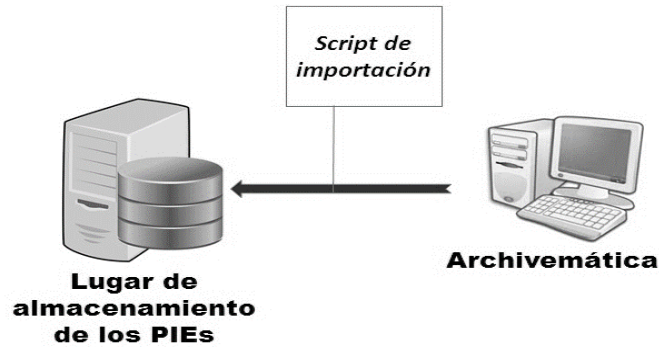


Figura 6: Esquema de migración.

Los paquetes de información fueron exportados por Dspace haciendo uso del plugin *packager*. Este permite exportar los documentos como paquete de información de archivo en forma comprimida conteniendo dos elementos:

- El documento digital exportado, el cual es renombrado por Dspace.
- Un archivo conteniendo los metadatos asociados al documento en formato METS.

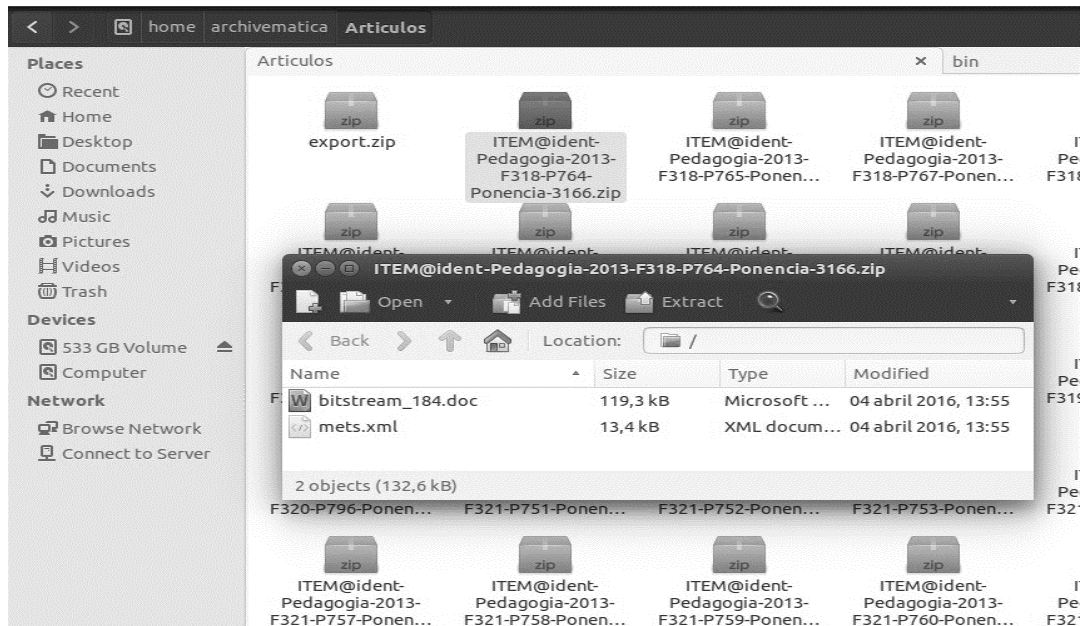


Figura 7: Composición de los paquetes de información de archivo generados por Dspace.

METS es uno de los formatos de codificación de metadatos más utilizados para preservar, ya que brinda información sobre cómo se crean y almacenan los archivos, los derechos de propiedad intelectual, información sobre la procedencia y la licencia, los cuales se deben tener en cuenta al realizar la preservación digital.

La importación de los documentos hacia Archivemática se realiza a través de una herramienta desarrollada por Artefactual, llamada Automation Tools, la cual es un conjunto de scripts en Python diseñados para automatizar el procesamiento de las transferencias en Archivemática ( GitHub, 2016). De esta herramienta se utilizó el script transfer.py, el cual es invocado desde el sistema operativo utilizando un script en lenguaje bash, quedando de la siguiente manera:

```
#!/bin/bash
cd /usr/lib/archivematica/automation-tools/
/usr/bin/python -m transfers.transfer --user <user> --api-key <apikey> --transfer-source
<transfer_source_uuid> --transfer-type 'dspace'
```

Los parámetros utilizados por el script transfer.py tienen la siguiente utilidad:

- --user: Usuario del dashboard con permisos a realizar la transferencia.
- --api-key: Llave que identifica al usuario.
- --transfer-source: ID de la ubicación de la fuente de transferencia para recuperar los documentos a transferir.
- --transfer-type: tipo de transferencia a iniciar, en este caso 'dspace'.

A continuación, se muestra un conjunto de imágenes correspondientes a una carga inicial realizada en Archivemática luego de ejecutado el script.

The screenshot shows the Archivemática web interface. At the top, there is a navigation bar with the Archivemática logo and menu items: Transferencia, Ingreso, Servicio de almacenamiento, Planificación de preservación, Acceso, and Administración. Below the navigation bar is a form for initiating a transfer. The form includes a dropdown menu for 'Tipo' (set to 'DSpace'), a text input for 'Nombre de transferencia', a text input for 'Número', a dropdown menu for the destination path (set to '/home/archivematica'), a 'Ver' button, and an 'Iniciar transferencia' button.

Below the form is a table showing the progress of the transfer. The table has three columns: 'Transferencia', 'Identificador UUID', and 'Comienzo'. The transfer is identified as 'ITEM@Ident-Pedagogia-2013-F318-P771-Ponencia.zip' and started on '2016-06-16 14:04'. The table lists various micro-services and tasks, with their completion status.

Transferencia	Identificador UUID	Comienzo
ITEM@Ident-Pedagogia-2013-F318-P771-Ponencia.zip		2016-06-16 14:04
▶ Micro-servicio: Crear SIP de Transferencia		
▶ Micro-servicio: Completar transferencia		
▶ Micro-servicio: Identificar archivo DSpace		
▶ Micro-servicio: Examinar contenido		
▼ Micro-servicio: Validación		
Tarea: Validar formato		Completado satisfactoriamente
▶ Micro-servicio: Caracterizar y extraer metadato		
▶ Micro-servicio: Actualizar METS.xml del documento		
▶ Micro-servicio: Extraer paquetes		
▶ Micro-servicio: Identificar formato del archivo		
▶ Micro-servicio: Limpiar nombre		
▶ Micro-servicio: Generar reporte de estructura de transferencia		
▼ Micro-servicio: Buscar virus		
Tarea: Análisis de virus		Completado satisfactoriamente
Tarea: Mover al directorio de procesamiento		Completado satisfactoriamente

Figura 8: Transferencia de los archivos exportados por Dspace hacia Archivemática

Paquete de información de envío	UUID	Comienzo
ITEM_ident-8036	5f906037-0848-45e6-8ddd-13b0cd36bf6d	2016-06-17 14:18
ITEM_ident-8035	58364ba6-4301-46ac-acb3-cf9d2bdca1a4	2016-06-17 14:14
ITEM_ident-8039	012a001f-ba70-49b0-a56c-8ecf143c8353	2016-06-17 14:07
A6	2cff84fe-41ee-41c0-8326-16cdb7441a09	2016-06-17 12:05
A5	d8c0a628-2a7f-47de-8c37-a3963bf54877	2016-06-17 12:02
A4	c29b6ec1-5a69-481c-8c69-ffea6406cdb8	2016-06-17 11:58
A3	01cdbc6e-2ce6-4b01-a437-e19ee6e28170	2016-06-17 11:55
A2	50d5a717-f4b7-4498-92ef-acbdc281ac8b	2016-06-17 11:46
A1	fb057795-ba83-42fa-8617-e237f0686036	2016-06-17 11:41

Figura 9: Ingreso de los paquetes de información a Archivematica.

Archivo(s)	AIP(s)
bitstream_4245.pdf 347a56b7-b69f-4fbf-aca3-41af1d4100de	G9 9b95d2dc-01f7-46e0-b565-113d00b22e6d (view raw)
bitstream_12385.pdf cc896a69-3742-4bc0-9c46-2ed4d30554e0	G8 a0685876-baae-4293-9180-79968c81ac4d (view raw)
bitstream_1780.pdf e1005a10-22b0-485b-b0aa-149873980705	G8 a0685876-baae-4293-9180-79968c81ac4d (view raw)
bitstream_4175.pdf a356b238-4bca-46f9-b14f-d97debde54d5	G8 a0685876-baae-4293-9180-79968c81ac4d (view raw)

Figura 10: Listado de los 771 archivos almacenados.

Como parte de la migración de datos, fueron ingresados en Archivematica un total de 6674 documentos digitales, correspondientes a la información científico técnica manejada por la Biblioteca de la UCI y gestionada a través del Repositorio Institucional.

## **2.3 Gestión de liberaciones y despliegue**

Una vez realizada la personalización de Archivematica para su adaptación a las necesidades de la Biblioteca de la UCI, el mismo puede ser desplegado en el entorno real en que será utilizado. Como parte de este proceso también se les ha brindado capacitación a los usuarios finales del sistema, y se les ha facilitado documentación del sistema para su auto-preparación.

### **2.3.1 Instalación y configuración de Archivematica en el entorno real**

Para realizar la instalación del sistema Archivematica primeramente es necesario tener un conjunto de requerimientos técnicos (Archivematica, 2015):

- Una computadora con Ubuntu LTS en su versión 12.04 o 14.04. Es recomendable utilizar la edición Server de 64 bits de Ubuntu 14.04.2.
- Como mínimo se requiere un procesador: Dual Core o i5 de 3ra generación.
- Un mínimo de memoria de 4Gb.
- El disco duro debe tener al menos 20Gb por encima del tamaño de la colección.
- De instalar Archivematica en varias máquinas, todas deben ser capaces de llegar a la otra por los siguientes puertos: http, mysqld, gearman, nfs, ssh.

Los paquetes de instalación de Archivematica están en un repositorio de internet, esto implica que se deben configurar un conjunto de parámetros en la máquina donde se va a instalar el sistema, entre ellos: las variables del entorno para poder acceder a internet desde la consola o terminal, se debe añadir el repositorio de Archivematica y de *ElasticSearch* a la lista de repositorios de confianza del sistema y se debe configurar el Python-pip<sup>8</sup> de la UCI para su utilización. Una vez configurado estos elementos se procede a realizar la instalación de los paquetes: *archivematica-storage-service*, *archivematica-mcp-server*, *archivematica-mcp-client*, *archivematica-dashboard* y *elasticsearch*. Luego de instalado los paquetes con sus dependencias se realiza la configuración del sistema. Dentro de los elementos a configurar se encuentra el servicio de almacenamiento y el *dashboard*.

Como parte de la configuración del sistema se establecieron las opciones principales del módulo administración, específicamente las asociadas al procesamiento, con las cuales se establecen los

---

<sup>8</sup> Python-pip es una herramienta que facilita la instalación de los paquetes de Python.

parámetros para realizar las transferencias. También se crearon los usuarios del sistema con sus respectivos roles y permisos, como se puede apreciar en la siguiente pantalla del sistema:

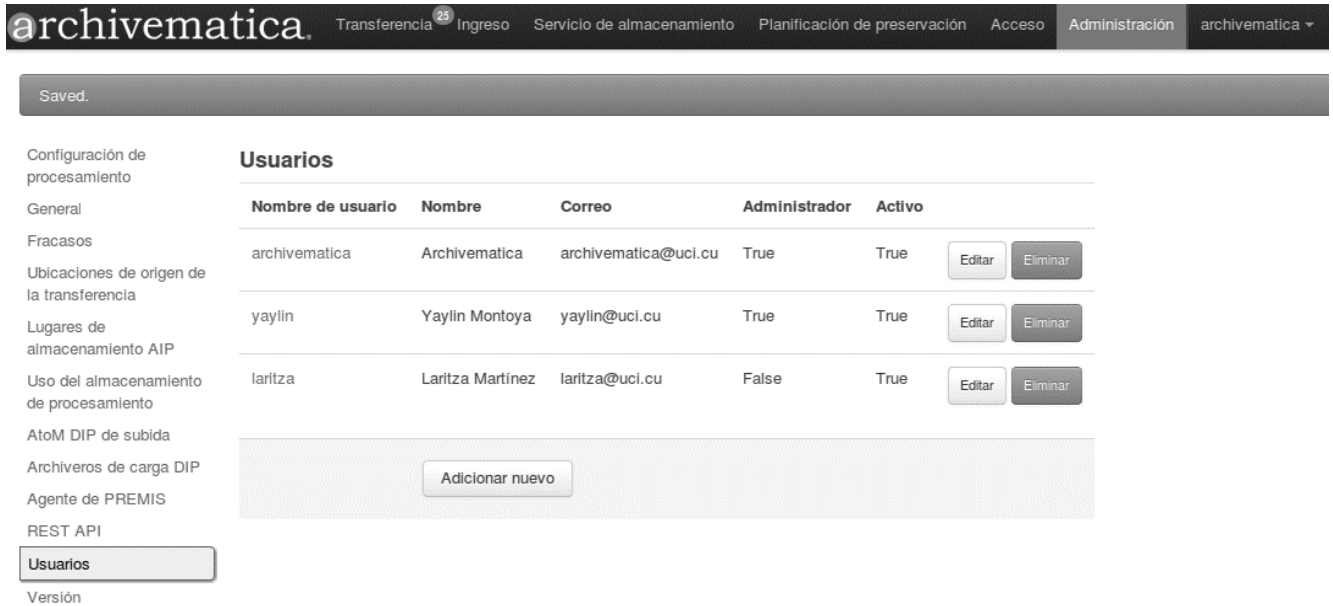


Figura 11: Listados de usuarios de Archivemática.

La etapa de despliegue merece gran importancia y cuidado, debido a que, si no se realiza adecuadamente, el éxito final se verá afectado sin importar cuán bien se halla trabajado en las otras fases. Su objetivo fundamental es la instalación y configuración del sistema, garantizando el correcto funcionamiento del mismo en el entorno real en donde será utilizado.

### 2.3.2 Capacitación

Para que el sistema implantado sea correctamente utilizado por los especialistas de la Biblioteca de la UCI, se les impartió una capacitación en el uso de Archivemática, la cual estuvo dividida en dos partes fundamentales, de acuerdo a los roles que desempeñarían en el sistema:

- A especialistas en ciencias de la información, cubriendo las funcionalidades básicas por cada uno de los módulos del sistema.
- A especialistas informáticos, con el fin de impartir los conocimientos básicos sobre la administración del sistema, así como las tecnologías más importantes utilizadas en su desarrollo.

Como parte de la capacitación se entregó a los participantes la documentación necesaria para el estudio de las funcionalidades del sistema de preservación instalado.



## 2.4 Validación y Pruebas de Servicios

Luego de realizado el despliegue del sistema en la biblioteca de la UCI, se le aplicaron pruebas de aceptación para validar que cumple con las expectativas del cliente.

### 2.4.1 Pruebas de aceptación

La validación se logra cuando el software funciona de acuerdo a las expectativas del cliente. Para comprobar que el sistema de preservación muestra los resultados esperados por el cliente, se realizaron los casos de prueba de aceptación. Las pruebas de aceptación son pruebas formales con respecto a las necesidades del usuario, conducidas para determinar si el sistema satisface o no el criterio de aceptación, y permite a los usuarios, clientes u otra entidad autorizada determinar si acepta o no el sistema. (STF, 2016)

A continuación, se muestran los casos de pruebas correspondientes a las funcionalidades Iniciar transferencia y Rechazar PIA, de los módulos Transferencia e Ingreso.

<b>Caso de prueba de aceptación</b>
<b>Nombre:</b> Iniciar transferencia.
<b>Descripción:</b> Prueba para la funcionalidad de Iniciar transferencia en el sistema de preservación.
<b>Condiciones de Ejecución:</b>
<ul style="list-style-type: none"> <li>✓ Seleccionar el tipo de transferencia Dspace.</li> <li>✓ Introducir el nombre de la transferencia.</li> <li>✓ Definir la ruta donde se encuentran los documentos a preservar.</li> </ul>
<b>Resultados esperados:</b>
<ul style="list-style-type: none"> <li>✓ El sistema comienza la transferencia, le asigna un identificador único y muestra las opciones para aceptar o rechazar dicha transferencia.</li> </ul>
<b>Evaluación de la prueba:</b> Prueba satisfactoria.

Tabla 6: Caso de Prueba para la funcionalidad Iniciar transferencia.

<b>Caso de prueba de aceptación</b>
<b>Nombre:</b> Rechazar PIA.
<b>Descripción:</b> Prueba para la funcionalidad Rechazar PIA en el ingreso.
<b>Condiciones de Ejecución:</b>
<ul style="list-style-type: none"> <li>✓ Debe existir el PIA.</li> </ul>
<b>Resultados esperados:</b>
<ul style="list-style-type: none"> <li>✓ El sistema muestra un mensaje de confirmación y las opciones Confirmar o Cancelar.</li> </ul>
<b>Evaluación de la prueba:</b> Prueba satisfactoria.

**Tabla 7: Caso de Prueba para la funcionalidad Rechazar PIA.**

Además de las validaciones formales realizadas a través de los casos de prueba de aceptación, se realizaron revisiones informales por parte de los usuarios finales y el cliente, en las cuales se identificaron las siguientes no conformidades:

- Se encontraron faltas de ortografía en algunas páginas del sistema.
- El micro-servicio sobre chequeo antivirus no funcionó correctamente en algunas transferencias.
- Palabras mal traducidas al idioma español.

Luego de corregidas la no conformidades por parte del equipo de desarrollo y la revisión posterior por el cliente, este último quedó satisfecho con las funcionalidades que implementa el sistema, emitiendo un acta de aceptación del sistema de preservación implantado, la cual puede ser encontrada en el Anexo 8.

Con las tareas descritas anteriormente se pudo verificar que se cumplieron las condiciones necesarias para que comience a funcionar el servicio de preservación digital en la Biblioteca de la UCI, a través del sistema Archivematica.

## 2.5 Beneficios de la propuesta

Dentro de los documentos digitales almacenados en la biblioteca de la UCI se encuentran los trabajos de diploma, tesis de maestría y doctorado, artículos científicos publicados en revistas y trabajos presentados en eventos, constituyendo los mismos recursos únicos y frutos del saber universitario. Muchos de estos documentos constituyen un patrimonio digno de protección y conservación, en beneficio de las generaciones actuales y futuras.

Son varias las amenazas por la que puede verse afectado un documento digital y esto puede ocasionar la pérdida de información de alto valor, por lo que la UNESCO alerta a sus Estados miembros a través de la *Carta para la preservación del Patrimonio digital* en el artículo 3, afirmando que: «El patrimonio digital del mundo corre el peligro de perderse para la posteridad. Contribuyen a ello, entre otros factores, la rápida obsolescencia de los equipos y programas informáticos que le dan vida, las incertidumbres existentes en torno a los recursos, la responsabilidad y los métodos para su mantenimiento y conservación y la falta de legislación que ampare estos procesos». También en el artículo 4 se expresa: “A menos que se haga frente a los peligros actuales, el patrimonio digital desaparecerá rápida e ineluctablemente”.

Con la implantación del sistema de preservación digital Archivematica se garantiza que los documentos digitales sean preservados a largo plazo sin perder las cualidades esenciales de disponibilidad, identidad, integridad y autenticidad. El sistema implantado hace uso de estrategias de preservación para mitigar las vulnerabilidades y amenazas a las que están sometidos los documentos digitales, ante el cambio rápido de los formatos aplica la migración, para garantizar que el documento continúe disponible y ser usado por los usuarios en el futuro.

Con el uso de este sistema se garantiza, además, que aun existiendo algún problema con el repositorio institucional de la biblioteca, exista la forma de que los documentos digitales no se pierdan ni queden obsoletos y puedan ser restaurados para su acceso por parte de los usuarios.

Después de terminado el capítulo se puede concluir que:

- Se logró implantar el sistema de preservación digital, internacionalizándose las interfaces más importantes del mismo.
- Se logró realizar la migración de los documentos digitales contenidos en el Repositorio institucional hacia el sistema de preservación.

### Capítulo 3: Validación de la propuesta de solución

Una manera de comprobar si lo que se ha hecho está correcto lo constituye la validación, este proceso se realiza con el objetivo de verificar la propuesta. En el presente capítulo y como parte de la validación se emplea el método Delphi. Se aplica este método basado en criterios de expertos, ya que la preservación digital cubre amplios períodos de tiempo, y no sería factible aplicar un método que de forma empírica logre verificar que el sistema puede realizar la preservación a largo plazo.

#### 3.1 Método Delphi

El método Delphi es un procedimiento eficaz y sistemático que tiene como objeto la recopilación de opiniones de expertos sobre un tema particular con el fin de incorporar dichos juicios en la configuración de un cuestionario y conseguir un consenso a través de la convergencia de las opiniones de expertos diseminados geográficamente (MIRA, 2010). La aplicación del método Delphi ejecuta las fases de formulación del problema, elección de expertos, elaboración y lanzamiento de los cuestionarios y por último desarrollo práctico y explotación de resultados.

En la presente investigación se utilizó este método, con el objetivo de validar que el sistema Archivematica posee las funcionalidades indispensables para realizar la preservación a largo plazo de la información científico-técnica de la Biblioteca de la UCI, manteniendo las cualidades de autenticidad y disponibilidad. A continuación, se exponen los resultados por cada fase.

#### 3.2 Fase 1: Formulación del problema

Se trata de una etapa fundamental en la realización de un Delphi. En un método de expertos, la importancia de definir con precisión el campo de investigación es muy grande, siendo preciso estar muy seguros de que los expertos reclutados y consultados poseen todos, la misma noción de este campo. La elaboración del cuestionario debe ser llevada a cabo según ciertas reglas: las preguntas deben ser precisas, cuantificables (versan por ejemplo sobre probabilidades de realización de hipótesis y/o acontecimientos, la mayoría de las veces sobre datos de realización de acontecimientos) e independientes (la supuesta realización de una de las cuestiones en una fecha determinada no influye sobre la realización de alguna otra cuestión. (CARDONA, 2004)

#### 3.3 Fase 2: Elección de expertos.

Se considerará un experto aquel individuo, grupo de personas u organizaciones que sean capaces de brindar valoraciones conclusivas sobre la propuesta, además de hacer recomendaciones con un

### Capítulo 3: Validación de la propuesta de solución

determinado coeficiente de competencia. Los mismos no solo deben ser conocedores del tema, sino que deben presentar gran diversidad en sus planteamientos. (Martínez, 2012)

Para este análisis se utiliza la metodología elaborada por el Comité Estatal para la Ciencia y la Técnica de Rusia, elaborado en 1971, para la determinación de la competencia de los expertos.

1. Confeccionar un listado inicial de personas posibles de cumplir los requisitos para ser expertos en la materia a trabajar, previamente consultada su disposición para participar.
2. Realizar una valoración sobre el nivel de experiencia que poseen, evaluando de esta forma los niveles de conocimientos que poseen sobre la materia. Se les realiza una primera pregunta para una autoevaluación de los niveles de información y argumentación que tienen sobre el tema en cuestión.
3. En esta pregunta se les pide que marquen con una X, en una escala creciente del 1 al 10, el valor que se corresponde con el grado de conocimiento o información que tienen sobre el tema a estudiar.

Expertos		1	2	3	4	5	6	7	8	9	10
1											
2											
3											

4. A partir de aquí se calcula fácilmente el Coeficiente de Conocimiento o Información (Kc), a través de la siguiente fórmula:

$$Kc = n(0,1) = \frac{n}{10}$$

Dónde:

Kc: Coeficiente de Conocimiento o Información

n: Rango seleccionado por el experto

5. Se realiza una segunda pregunta que permite valorar un grupo de aspectos que influyen sobre el nivel de argumentación o fundamentación del tema a estudiar.

Fuentes de argumentación o fundamentación		Alto	Medio	Bajo
1	Su experiencia teórica			
2	Su experiencia práctica			

### Capítulo 3: Validación de la propuesta de solución

3	Bibliografía nacional consultada			
4	Bibliografía internacional consultada			
5	Su conocimiento del estado del problema			
6	Su intuición			

6. Aquí se determinan los aspectos de mayor influencia. A partir de estos valores reflejados por cada experto en la tabla se contrastan con los valores de una tabla patrón:

Fuentes de argumentación o fundamentación		Alto	Medio	Bajo
1	Su experiencia teórica	0.30	0.20	0.10
2	Su experiencia práctica	0.50	0.40	0.20
3	Bibliografía nacional consultada	0.05	0.05	0.05
4	Bibliografía internacional consultada	0.05	0.05	0.05
5	Su conocimiento del estado del problema	0.05	0.05	0.05
6	Su intuición	0.05	0.05	0.05

7. Los aspectos que influyen sobre el nivel de argumentación o fundamentación del tema a estudiar permiten calcular el Coeficiente de Argumentación ( $K_a$ ) de cada experto:

$$K_a = \sum ni = (n_1 + n_2 + n_3 + n_4 + n_5 + n_6)$$

Dónde:

$K_a$ : Coeficiente de Argumentación

$ni$ : Valor correspondiente a la fuente de argumentación  $i$  (1 hasta 6)

Una vez obtenido los valores del Coeficiente de Conocimiento ( $K_c$ ) y el Coeficiente de Argumentación ( $K_a$ ) se procede a obtener el valor del Coeficiente de Competencia ( $K$ ) que finalmente es el coeficiente que determina en realidad que experto se toma en consideración para trabajar en esta investigación. Este coeficiente ( $K$ ) se calcula de la siguiente forma:

$$K = 0.5(K_c + K_a) = \frac{(K_c + K_a)}{2}$$

Dónde:

$K$ : Coeficiente de Competencia

Kc: Coeficiente de Conocimiento

Ka: Coeficiente de Argumentación

8. Posteriormente obtenido los resultados se analizan de la manera siguiente:

$0,8 \leq K \leq 1,0$  Coeficiente de Competencia Alto

$0,5 \leq K < 0,8$  Coeficiente de Competencia Medio

$K < 0,5$  Coeficiente de Competencia Bajo

Se debe utilizar para la consulta a expertos de competencia alta, no obstante, puede se puede valorar utilizar expertos de competencia media en caso de que el coeficiente de competencia promedio de los posibles expertos sean alto, pero nunca se utilizará expertos de competencia baja.

Se tuvo en cuenta como criterio principal para la selección de los posibles expertos a participar en la validación de este trabajo aquellos profesionales cuyas competencias y conocimientos estuviesen relacionados a las ciencias de la información, ya que la preservación digital a largo plazo se enmarca en el área de estudio antes mencionada. De ahí que la composición final de los expertos cuente en menor medida con ingenieros informáticos, ya que la preservación digital no es un tema de estudio afín a los graduados de esta especialidad.

Luego de definido los requisitos que deben cumplir los encuestados para ser seleccionados como expertos, se encuestaron cuatro posibles expertos de la Biblioteca Nacional: "José Martí" y cinco de la UCI. Una vez aplicada la encuesta se procesaron los datos y se realizan los cálculos pertinentes, los resultados se muestran a continuación.

No	Nombre y Apellidos	K	Competencia	Seleccionado
1	Luis Carlos Alvarez Fernández	0,8	Alto	Si
2	Luis Bravo Martínez	0,35	Bajo	No
3	Dayaris Flores Álvarez	0,3	Bajo	No
4	Laritza Martínez Negrín	0,75	Medio	Si
5	Silvana Pérez Zappino	0,6	Medio	Si
6	Sonia Núñez Amaro	0,5	Medio	Si
7	Jorge Dayán Aguiar Cedeño	0,7	Medio	Si
8	Leandro Tabares Martín	0,7	Medio	Si
9	Olga Lidia Martínez Acosta	0,9	Alto	Si

Tabla 8: Resultado de aplicar los criterios a los posibles expertos.

Solamente los candidatos con Competencia media y alto fueron considerados expertos, lo que resulto en un total de 7, que es la cantidad mínima exigida por este método.

#### 3.4 Fase 3: Elaboración y lanzamiento de los cuestionarios

Luego de realizar la selección de los expertos, se puede aplicar el método Delphi para el procesamiento de los criterios y opiniones recogidas mediante un cuestionario. El cuestionario se elaboró de manera que facilite la respuesta por parte de los expertos.

Para elaborar el cuestionario se utilizó la propuesta del Catálogo de criterios para Repositorios Digitales Confiables (*Catalogue of Criteria for Trusted Digital Repositories*), versión 2, creado por el grupo NESTOR (*Network of Expertise in Long-Term Storage of Digital Resources*), el cual está dirigido principalmente a las organizaciones de la memoria (archivos, bibliotecas, museos) y sirve como un manual para la elaboración, planificación e implementación de un repositorio digital de confianza a largo plazo. Este catálogo identifica criterios que permiten evaluar un repositorio digital en los niveles organizacional y técnico. Los criterios fueron definidos en estrecha colaboración con un amplio rango de diferentes organizaciones de la memoria, productores de información, expertos y otras partes interesadas. Esta propuesta asegura un alto grado de validez universal para uso práctico cotidiano y también una amplia aceptación de los resultados. (NESTOR Working Group, 2009)

El cuestionario fue creado de manera que las respuestas fueran categorizadas entre Muy Adecuado, Bastante Adecuado, Adecuado, Poco Adecuado, No Adecuado. En el Anexo 2 se muestra el cuestionario de validación utilizado en la investigación.

#### 3.5 Fase 4: Desarrollo práctico y explotación de resultados

Los expertos recibieron un resumen de la propuesta como documentación para contestar la encuesta, la cual consta de un total de diecisiete preguntas. Se realizó una ronda de encuesta y luego se procedió al análisis de los resultados.

Luego de tener los datos computados se procede a la ejecución de los siguientes pasos para la obtención de los resultados esperados.

**Paso 1:** construcción de la tabla de frecuencia observada, suma de las valoraciones por aspectos de cada experto. (Ver Anexo 3)

**Paso 2:** construcción de la tabla de frecuencias acumuladas. En esta el número de la fila, se obtiene sumando cada valor con el anterior. (Ver Anexo 4)



### Capítulo 3: Validación de la propuesta de solución

**Paso 3:** se construye la tabla de frecuencias relativas acumuladas, estos valores se obtienen de dividir los valores numéricos de la tabla anterior entre el número total de expertos. (Ver Anexo 5)

**Paso 4:** Asignación a partir de la tabla de Z de la distribución normal, del valor de la imagen que corresponde a cada frecuencia acumulativa relativa obtenida. (Ver Anexo 6)

**Paso 5:** Obtención de los puntos a través del cálculo de N-P, donde:

$$N = \frac{\text{Sumatoria de la Suma por Aspectos}}{(\text{No. de Rangos de Valoración} \times \text{No. de Aspectos})}$$

P= Promedio por Aspectos

**Paso 6:** Se divide la recta por categorías a partir de los Puntos de Corte y se ubican los puntos N-P para determinar la categoría de cada aspecto.

$$N = \frac{\text{Sumatoria Rasgos de Valoración}}{(\text{No. de Aspectos a Evaluar})}$$

**Paso 7:** Análisis de la información obtenida.

Con estos resultados podemos otorgar los respectivos rangos de valoración a cada aspecto analizado, es decir, Muy Adecuado, Bastante Adecuado, Adecuado, Poco Adecuado e Inadecuado o sencillamente otorgarle un grado de jerarquía del 1 al 5 respectivamente.

Si hay variación respecto a lo que plantearon inicialmente, se vuelven a procesar los nuevos rangos de valoración con el mismo procedimiento y se consultan las veces que sea necesario hasta que no haya más modificaciones en los criterios. Se realizan cuantas vueltas sean necesarias hasta que los expertos decidan no variar más su opinión.

Los puntos de corte se utilizan para determinar la categoría o grado de adecuación de cada criterio según la opinión de los expertos consultados. Con ellos se opera del modo que expresa la tabla 5:

	Inadecuado	Poco Adecuado	Adecuado	Bastante Adecuado	Muy Adecuado
Punto/ Corte	–	–	–	-3,09 < N-P <= -1,16436	N-P > -1,16436

Tabla 9: Puntos de cortes.

Luego de realizado todos los pasos se puede concluir que:

- La evaluación final de los expertos a los 17 criterios fue de Muy Adecuada. (Ver Anexo 7)

- El 100% de los expertos coinciden en la utilidad que tiene el sistema Archivematica para la preservación digital a largo plazo de la información científico-técnica.

#### 3.6 Establecimiento de la concordancia entre los Expertos

Para que la propuesta tenga mayor validez debe existir un excelente acuerdo entre los expertos, por tal razón es que se calcula el Coeficiente de Concordancia de Kendall que posibilita comprobar el grado de coincidencia de las valoraciones realizadas por los expertos. El Coeficiente de Concordancia de Kendall (W) se obtiene aplicando la siguiente fórmula:

$$W = \frac{12 * S}{K^2(N^3 - N)}$$

Donde la suma de los cuadrados de las desviaciones de la media (S) se obtiene de la sumatoria de los rangos (Sj) entre N, siendo N el total de aspectos a evaluar (los aspectos son las preguntas del cuestionario), K es el número total de expertos. En el Anexo 8 se muestran los cálculos realizados para determinar la concordancia de los expertos.

K es el número de expertos que intervienen en el proceso de validación, por lo tanto, K=7.

N, cantidad de aspectos a validar, en este caso N=17.

Sj, suma de los rangos asignados a cada pregunta por parte de los expertos.

La media de los rangos se determina a través de la fórmula:

$$\bar{S}_j = \frac{\sum_{j=1}^N S_j}{N}$$

Obteniendo el valor de: 33,4705882352941

Para resumir en varios pasos lo que se ha planteado anteriormente y lo que se hará a continuación, sirve como guía la metodología establecida a continuación:

- Determinación de la suma de los valores numéricos asignados a cada aspecto a evaluar, según la apreciación del experto (Rj).
- Determinación del valor medio de las Rj, dada por la sumatoria de los Rj entre N, siendo N el total de aspectos a evaluar.
- Determinación de la desviación media (S), dada por la diferencia entre cada Rj y el valor de la media.
- Determinación de la suma de los cuadrados de las desviaciones medias.
- Determinación del cuadrado del número total de expertos (K).
- Determinación del cubo del número total de aspectos a evaluar(N).

### Capítulo 3: Validación de la propuesta de solución

- Determinación de la diferencia entre el cubo de N y N y su multiplicación por el cuadrado de K.
- Determinación del estadígrafo.

Para determinar la desviación media se utiliza la siguiente ecuación:

$$\sum_{j=1}^N (S_j - \bar{S})^2$$

Después de realizado el cálculo:  $S=26$ , se puede calcular el coeficiente Kendall y se obtiene a través de la ecuación:

$$W = \frac{12 * S}{K^2(N^3 - N)}$$

Sustituyendo los valores en la ecuación:

$$W = \frac{12 * S}{K^2(N^3 - N)} = 0.00130052020808323$$

El coeficiente de Kendall obtenido permite calcular el Chi cuadrado real, el cual tiene el objetivo de medir si existe o no concordancia entre los expertos y se obtiene a través de la fórmula:

$$x^2 = K(N - 1)W$$

$$x^2 = 7(17 - 1)0.00130052020808323$$

$$x^2 = 0.1456582633053184$$

Este Chi-Cuadrado se compara con el de las tablas estadísticas con una probabilidad de error de 0,01 y N (número de clases). Si el Chi-Cuadrado real es menor que el Chi-Cuadrado de la tabla entonces hay concordancia:

$$x^2_{real} < x^2(\alpha, N - 1)$$

En este caso el valor de N es 2, ya que de las 5 clases que se tenían inicialmente (Muy Adecuado, Bastante Adecuado, Adecuado, Poco Adecuado e Inadecuado), solamente se utilizaron 2 (Muy Adecuado y Bastante Adecuado).

$$x^2_{real} < x^2(0.01, 2 - 1)$$

$$0.14 < 6,6349$$

Por lo tanto, se puede concluir que hay concordancia entre los expertos.

### **Capítulo 3: Validación de la propuesta de solución**

---

Con la validación de la propuesta queda evidenciado el cumplimiento de los objetivos propuestos y que los expertos participantes ratificaron esta afirmación con su valoración.

### Conclusiones generales

Con la realización del presente trabajo de diploma se han obtenido las siguientes conclusiones:

- El estudio de los principales elementos teóricos y los sistemas de preservación existentes en el mundo permite concluir que la preservación digital es un área de estudio interdisciplinar, con una fuerte integración de las ciencias de la información y la informática. Todavía no existe una solución definitiva al problema de la preservación digital.
- Luego de analizado el contexto organizacional se determinó que la preservación digital aún no constituye una prioridad, lo cual se refleja en la ausencia de procesos para tal fin en la Biblioteca de la UCI.
- La implantación de Archivematica fue realizada utilizando una metodología coherente con el objetivo de la investigación. El uso de este sistema en la Biblioteca de la UCI, permitirá reducir los riesgos a los que está expuesta la información científico-técnica, al aplicar estrategias y herramientas adecuadas para su preservación digital a largo plazo.
- La validación del sistema a través del método Delphi, permitió comprobar que Archivematica posee las funcionalidades adecuadas para realizar la preservación digital a largo plazo de la información científico-técnica de la Biblioteca de la UCI, manteniendo sus cualidades de autenticidad y disponibilidad.
- El sistema implantado, al ser de libre distribución, constituye una alternativa a variantes propietarias, con lo cual se reducen gastos asociados al pago de licencias y actualizaciones de software.

### Recomendaciones

- Se recomienda integrar la solución propuesta al repositorio institucional, utilizando alguno de los protocolos proporcionados por este, ya sea OAI-PMH o OAI-ORE. Esta integración debe realizarse con actualizaciones mejoradas de Dspace ya que la actual no los soporta correctamente.
- Realizar los trámites para la instalación de Archivematica en el nodo central de la UCI, para un mejor rendimiento y disponibilidad del sistema.
- Realizar una investigación para implementar a Archivematica nuevas funcionalidades asociadas al almacenamiento distribuido de los documentos.
- Configurar Archivematica para la ejecución de los micro-servicios de forma distribuida usando varias computadoras, con lo cual se mejoraría el rendimiento del sistema.

### Referencias Bibliográficas

**GitHub. 2016.** GitHub. [En línea] 2016. <https://github.com/artefactual/fixity>.

**ITIL & IT SERVICE MANAGEMENT BOOKSHOP. 2014.** [En línea] 2014. <http://www.itil.org.uk/index.htm>.

**Lamarca Lapuente M. 2013.** *DOCUMENTO/HIPERTEXTO. Hipertexto: El nuevo concepto de documento en la cultura de la imagen.* 2013.

**NGINX . 2016.** NGINX. [En línea] NGINX Inc, 2016. <https://www.nginx.com/resources/wiki/#.CA94107-1-800-915-9122>.

**ACUP. 2012.** *e-Arxiu, nou servei d'administració electrònica per a les universitats catalanes.* 2012.

**Alegsa, L. 2010.** *DICCIONARIO DE INFORMATICA Y TECNOLOGIA.* 2010.

**Archivematica. 2015.** Archivematica. [En línea] 2015. [Citado el: 27 de 4 de 2016.] <https://www.archivematica.org/en/>.

**Artefactual. 2016.** Artefactual. [En línea] 2016. <https://www.artefactual.com/services/maintenance/>.

**Barateiro, J., Antunes, G., & Borbinha, J. 2010.** *Addressing Digital Preservation: Proposals for New Perspectives.* 2010.

**BARRUECO, JOSÉ MANUEL. 2011.** *Preservación y conservación de documentos digitales.* 2011.

**Barrueco, José Manuel y Coll, Imma Subirats. 2003.** *Open archives initiative. Protocol for metadata harvesting (OAI-PMH): descripción, funciones y aplicaciones de un protocolo.* 2003.

**Betanzos, Genaro Antonio León. 2015.** *Manual para administración y uso de DSpace.* 2015.

**Biblioteca Nacional de Australia. 2003.** *Directrices para la preservación del patrimonio digital.* 2003.

**Blanco, Y. 2011.** *Módulo de Conservación del Sistema de Gestión de Documentos de Archivos ArchiVenHIS (Trabajo de diploma).* Universidad de las Ciencias Informáticas. La Habana : s.n., 2011.

**Boudrez, F. 2007.** Digital signatures and electronic records. *Archival Science.* 2007.

**BRAVO, PÍA RAMÍREZ y JAURÈS, FELIPE DONOSO. 2006.** *METODOLOGÍA ITIL. Descripción, Funcionamiento y Aplicaciones.* Santiago de Chile : s.n., 2006.

**Campillo, I. 2010.** *Sistema de Gestión Integral de Documentos de archivo para empresas de la construcción del territorio de Camagüey (Tesis doctoral).* UNIVERSIDAD DE LA HABANA. 2010.

**CARDONA, NATALIA MARGARITA CASTAÑO. 2004.** *UNIVERSIDAD-EMPRESA LA OPORTUNIDAD QUE TIENE COLOMBIA PARA INNOVAR EN TECNOLOGÍAS INFORMÁTICAS.* Departamento de Ingeniería de Sistemas, UNIVERSIDAD JAVERIANA . SANTAFÉ DE BOGOTA : s.n., 2004.

**Carrazana, Castro E. 2015.** *Propuesta de requisitos funcionales para la preservación a largo plazo de información digital Universidad de las Ciencias Informáticas.* Universidad de la Habana-Facultad de comunicación. La Habana : s.n., 2015.

- CCSDS. 2012.** *REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS)*. Washington : MAGENTA BOOK, 2012.
- Conway, P. 1996.** *Preservation in the Digital World*. Yale University Library. 1996.
- Dspace. 2016.** Dspace. [En línea] 2016. <http://www.dspace.org/>.
- Duranti, L. 2001.** *Concepts, Principles, and Methods for the Management of Electronic Records*. The Information Society. 2001.
- Duranti, L. 1995.** *Reliability and Authenticity: The Concepts and Their Implications*. *Archivaria*. 1995. págs. 5-10.
- Ewart, John. 2013.** *Parallel Processing with Gearman*. s.l. : Packt Publishing, 2013.
- GARZÓN PÉREZ, M. 2010.** *SISTEMAS GESTORES DE BASES DE DATOS*. 2010.
- Gilliland, A. 2005.** *Ensayo de discusión sobre la naturaleza y el rol de los metadatos en la creación de documentos fiables y en la conservación de documentos auténticos en sistema electrónicos*. 2005.
- Gómez Labrador, R. septiembre de 2005.** *TIPOS DE LICENCIAS DE SOFTWARE*. septiembre de 2005.
- González Duque, R. 2011.** *Python para todos*. 2011.
- GONZÁLEZ, GRENDA P. AGUIRRE. 2014.** *Módulo de Administración e Integración*. 2014.
- González, Lianet Cabrera. 2012.** *EXTENSIÓN DE VISUAL PARADIGM FOR UML PARA EL DESARROLLO DIRIGIDO POR MODELOS DE APLICACIONES DE GESTIÓN DE INFORMACIÓN*. 2012.
- Hoeven, J., & Wijngaarden, H. 2005.** *Modular emulation as a long-term preservation strategy for digital objects*. the National Library of the Netherlands. s.l. : Koninklijke Bibliotheek, 2005.
- Instituto de Economía Aplicada a la Empresa. 2001.** *APLICACIÓN DEL MÉTODO DELPHI EN LA ELABORACIÓN DE LA TABLA SIMÉTRICA DE LAS TABLAS INPUT-OUTPUT* . 2001.
- Internacional council on archives. 2010.** . *Digital preservation [ICA Terminology Database]*. 2010.
- InterPARES 2 Project. 2008.** The InterPARES 2 Project Glossary. *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*. [En línea] 2008. [Citado el: 17 de 2 de 2016.] [http://www.interpares.org/display\\_file.cfm?doc=ip2\\_book\\_glossary.pdf](http://www.interpares.org/display_file.cfm?doc=ip2_book_glossary.pdf).
- InterPARES. 2013.** The International Research on Permanent Authentic Records in Electronic Systems. [En línea] 2013. [Citado el: 17 de 2 de 2016.] <http://interpares.org/>.
- ISO. 2011.** *ISO 15489-1:2001 - ISO/TR 15489-2 Information and Documentation - Records Management*. 2011.
- Kabir, Mohammed J. 2003.** *Servidor Apache 2*. 2003. 978-84-415-1468-3 84-415-1468-2.
- Keefer A, Gallart N. 2014.** *La preservación de recursos digitales: El reto para las bibliotecas del siglo XXI*. 2014.



- Kempton, S.** 2016. [En línea] 2016. [http://wiki.en.it-processmaps.com/index.php/ITIL\\_Service\\_Transition](http://wiki.en.it-processmaps.com/index.php/ITIL_Service_Transition).
- Lavoie, B., & Gartner, R.** 2013. *Preservation Metadata (2nd edition)*. 2013.
- Library of Congress.** 2015. *Sustainability of Digital Formats Planning for Library of Congress Collections*. 2015.
- Lupovici, C., & Masanès, J.** 2000. *Metadata for the Long Term Preservation of Electronic Publications*. Biblioteca Nacional de Francia. 2000.
- MacNeil, H.** 2000. *Trusting records: legal, historical and diplomatic perspectives*. Kluwer Academic Publishers. 2000.
- Martínez, Daymi Vega.** 2012. *Estrategia Formativa de Medio Ambiente del currícul Industrial*. Matanzas : s.n., 2012.
- Mena Mugica, M. M.** 2006. *Propuesta de requisitos funcionales para la gestión de documentos archivísticos electrónicos en la administración central del estado cubano (Tesis doctoral)*. Universidad de La Habana. Ciudad de La Habana. : s.n., 2006.
- Méndez, Eva y Senso, José A.** 2004. Introducción a los Metadatos: Estandares y aplicación. *Uso del Dublin Core (DCMI). ISO 15836-2003*. [En línea] Unidad de Autoformación. SEDIC , 2004. [Citado el: 31 de 3 de 2016.] <http://www.sedic.es/autoformacion/metadatos/tema7.htm>.
- MENGUAL, S.** 2011. *La importancia percibida por el profesorado y el alumnado sobre la inclusión de la competencia digital en educación Superior, Alicante, Departamento de Didáctica General y Didácticas específicas de la Facultad de Alicante*. 2011.
- MIRA, JOSEFA E. BLASCO.** 2010. *VALIDACIÓN MEDIANTE MÉTODO DELPHI DE UN CUESTIONARIO PARA CONOCER LAS EXPERIENCIAS E INTERÉS HACIA LAS ACTIVIDADES ACUÁTICAS CON ESPECIAL ATENCIÓN AL WINDSURF*. 2010.
- Morrissey, S. M., Meyer, J., & Bhattarai, S.** 2010. *Portico: A Case Study in the Use of the Journal Archiving and Interchange Tag Set for the Long Term Preservation of Scholarly Journals*. 2010.
- NESTOR Working Group.** 2009. *Catalogue of Criteria for Trusted Digital Repositories- Version 2*. 2009. 0008-2010030806.
- Ortiz Ramirez, A.** 2010. *Python como primer lenguaje de programación*. 2010.
- Preservica.** 2015. Achieving a Step Change in Digital Preservation Capability. [En línea] 2015. <http://preservica.com/resource/essential-guide-achieving-step-change-digital-preservation->
- Prieto A., Lloris A., & Torres J.** 2002. *Introducción a la Informática*. 2002.
- Quisbert, H.** 2008. *On Long-term Digital Preservation Information Systems (Tesis doctoral)*. Luleå University of Technology. 2008.
- Ras, M.** 2009. *The KB e-Depot: Building and Managing a Safe Place for e-Journals*. *Liber Quarterly*. 2009.

- Rauber A., Kaiser M., Guenther R., Constantopoulos P. 2010.** *Proceedings of the 7th International Conference on Preservation of Digital Objects.* 2010.
- Real Academia Española. 2016.** Real Academia Española. [En línea] 2016. <http://www.rae.es/>.
- Rosenthal D., Robertson T., Lipkis T., Reich V., & Morabito S. 2005.** *Requirements for Digital Preservation Systems: A Bottom-Up Approach.* *D-Lib Magazine.* 2005.
- Rothenberg, J. 1999.** *Ensuring the Longevity of Digital Information.* Santa Monica : s.n., 1999.
- SANNER, M. F. 1999.** *PYTHON: A PROGRAMMING LANGUAGE FOR SOFTWARE INTEGRATION AND DEVELOPMENT.* 1999.
- Seadle, M. 2013.** *EU digital preservation projects.* *En Digital Archiving.* Berlin : Humbolt-Universitat zu, 2013.
- Sherry A. Southerland, Vivian L. Gadsden, Carolyn D. Herrington. 2014.** *Editors' Introduction: What Should Count as Quality Education Research? Continuing the Discussion.* 2014.
- Song S. & JaJa J. 2009.** *Techniques to audit and certify the long-term integrity of digital archives.* *International Journal on Digital Libraries.* 2009.
- STF. 2016.** [En línea] 2016. <http://softwaretestingfundamentals.com/acceptance-testing/>.
- Suehring, Steve. 2001.** *MySQL Bible.* 2001.
- Ureña Almagro, C. 2011-2012.** *Lenguajes de Programación.* 2011-2012.
- Vazquez, Y. 2013.** *Módulo de Conservación para el Sistema de Gestión de Documentos Históricos Dextra (Trabajo de diploma).* Universidad de las Ciencias Informáticas. La Habana : s.n., 2013.
- Vermaaten, S., Lavoie, B., & Caplan, P. 2012.** *Identifying Threats to Successful Digital Preservation: the SPOT Model for Risk Assessment.* *D-Lib Magazine.* 2012.
- Vernooy, M. 2009.** *Enhanced Publications: Linking Publications and Research Data in Digital Repositories.* Amsterdam: Amsterdam University Press. 2009.
- Wang X., & Feng D. 2004.** *Collisions for Hash Functions MD4, MD5, HAVAL-128 and RIPEMD.* *International Association for Cryptologic Research.* 2004.

## Anexo 1

### Encuesta de autoevaluación para la selección de los expertos

**Compañero (a):**

Como parte importante y final de la presente investigación se pretende realizar una validación de la propuesta tomando como base los criterios de un grupo de expertos en el tema. Para esto es necesario conocer el grado de dominio que usted posee acerca de la Preservación digital. Por lo que se le pide llenar los datos de la siguiente tabla.

<b>Nombre y apellidos</b>	
<b>Centro de Trabajo</b>	
<b>Años de experiencia</b>	
<b>Especialidad</b>	

**1- Marque con una X el grado de conocimiento que usted posee del tema de Preservación digital:**

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>

**2- Marque con una X las fuentes que le han servido a usted para adquirir el conocimiento sobre el tema.**

<b>No</b>	<b>Fuentes de argumentación</b>	<b>Grado</b>		
		<b>Alto</b>	<b>Medio</b>	<b>Bajo</b>
<b>1</b>	Su experiencia teórica			
<b>2</b>	Su experiencia práctica			
<b>3</b>	Bibliografía nacional consultada			
<b>4</b>	Bibliografía internacional consultada			
<b>5</b>	Su conocimiento del estado del problema			
<b>6</b>	Su intuición			

## Anexo 2

### Encuesta para la validación de la propuesta.

Estimado experto (a):

La siguiente encuesta tiene como objetivo obtener su evaluación acerca de la propuesta de solución. Para lograr este objetivo se le proponen un conjunto de preguntas que permiten medir si el sistema implantado en la Biblioteca de la UCI, posee las funcionalidades necesarias para preservar a largo plazo la información digital. De antemano se le asegura el anonimato, de tal forma que usted no podrá ser identificado a través de las respuestas que brinde. Valore el grado de factibilidad de cada pregunta de acuerdo con la siguiente escala:

Muy adecuado (MA), Bastante Adecuado (BA), Adecuado (A), Poco Adecuado (PA) e Inadecuado (I).

¡Gracias por su tiempo!

Preguntas	Criterios				
	MA	BA	A	PA	I
1. Archivemática garantiza la integridad de los documentos durante el ingreso.					
2. Archivemática implementa las funciones de almacenamiento de archivos requeridas para chequear y mantener la integridad de los objetos digitales.					
3. Archivemática asegura la integridad de los objetos digitales durante el acceso.					
4. Archivemática garantiza la autenticidad de los objetos digitales recibidos durante el ingreso.					
5. Archivemática garantiza la autenticidad de los objetos digitales almacenados.					
6. Archivemática garantiza la autenticidad de los paquetes de acceso.					
7. Archivemática define los formatos a utilizar para los paquetes de información de ingreso.					
8. Archivemática define los formatos a utilizar para los paquetes de información de archivo.					
9. Como parte del proceso de ingreso, Archivemática se encarga de transformar los paquetes de información de ingreso en paquetes de información de archivo.					

10. Archivemática implementa estrategias para la preservación a largo plazo de los paquetes de información de archivo.					
11. Archivemática asegura la transformación de paquetes de información de archivo en paquetes de información de diseminación.					
12. Archivemática identifica de forma exclusiva y persistente los objetos digitales.					
13. Archivemática registra metadatos adecuados para la descripción estructural de los objetos digitales.					
14. Se registran metadatos adecuados para documentar todos los cambios realizados por el sistema a los objetos digitales.					
15. Archivemática adquiere metadatos adecuados para la descripción técnica de los objetos digitales.					
16. El sistema adquiere metadatos adecuados para registrar los derechos y condiciones de uso de los documentos digitales.					
17. ¿Considera usted que el sistema implantado en la Biblioteca de la UCI, posee las funcionalidades adecuadas para realizar la preservación a largo plazo de la información digital?					
Expresar otros criterios o recomendaciones que pudieran servir para perfeccionar la propuesta:					

**Anexo 2**

Tabla de frecuencia Observada

<b>Criterio</b>	<b>Inadecuado</b>	<b>Poco adecuado</b>	<b>Adecuado</b>	<b>Bastante adecuado</b>	<b>Muy adecuado</b>
<b>1</b>	0	0	0	1	6
<b>2</b>	0	0	0	0	7
<b>3</b>	0	0	0	3	4
<b>4</b>	0	0	0	2	5
<b>5</b>	0	0	0	0	7
<b>6</b>	0	0	0	0	7
<b>7</b>	0	0	0	3	4
<b>8</b>	0	0	0	3	4
<b>9</b>	0	0	0	1	6
<b>10</b>	0	0	0	0	7
<b>11</b>	0	0	0	2	5
<b>12</b>	0	0	0	1	6
<b>13</b>	0	0	0	3	4
<b>14</b>	0	0	0	2	5
<b>15</b>	0	0	0	1	6
<b>16</b>	0	0	0	3	4
<b>17</b>	0	0	0	2	5

**Anexo 4**

Tabla de Frecuencia Absoluta.

<b>Criterio</b>	<b>Inadecuado</b>	<b>Poco adecuado</b>	<b>Adecuado</b>	<b>Bastante adecuado</b>	<b>Muy adecuado</b>
1	0	0	0	1	7
2	0	0	0	0	7
2	0	0	0	3	7
4	0	0	0	2	7
5	0	0	0	0	7
6	0	0	0	0	7
7	0	0	0	3	7
8	0	0	0	3	7
9	0	0	0	1	7
10	0	0	0	0	7
11	0	0	0	2	7
12	0	0	0	1	7
13	0	0	0	3	7
14	0	0	0	2	7
15	0	0	0	1	7
16	0	0	0	3	7
17	0	0	0	2	7

## Anexo 5

Tabla de frecuencia Acumulada Relativa

criterio	Inadecuado	Poco adecuado	Adecuado	Bastante adecuado
1	0	0	0	0,142857149243355
2	0	0	0	0
2	0	0	0	0,428571432828903
4	0	0	0	0,28571429848671
5	0	0	0	0
6	0	0	0	0
7	0	0	0	0,428571432828903
8	0	0	0	0,428571432828903
9	0	0	0	0,142857149243355
10	0	0	0	0
11	0	0	0	0,28571429848671
12	0	0	0	0,142857149243355
13	0	0	0	0,428571432828903
14	0	0	0	0,28571429848671
15	0	0	0	0,142857149243355
16	0	0	0	0,428571432828903
17	0	0	0	0,28571429848671



## Anexo 6

Tabla de distribución normal Inversa correspondiente a cada Frecuencia Acumulativa Relativa.

<b>Criterio</b>	<b>Inadecuado</b>	<b>Poco adecuado</b>	<b>Adecuado</b>	<b>Bastante adecuado</b>
<b>1</b>	-3,09	-3,09	-3,09	-1,06757049672431
<b>2</b>	-3,09	-3,09	-3,09	-3,09
<b>2</b>	-3,09	-3,09	-3,09	-0,180012358854845
<b>4</b>	-3,09	-3,09	-3,09	-0,565948784978835
<b>5</b>	-3,09	-3,09	-3,09	-3,09
<b>6</b>	-3,09	-3,09	-3,09	-3,09
<b>7</b>	-3,09	-3,09	-3,09	-0,180012358854845
<b>8</b>	-3,09	-3,09	-3,09	-0,180012358854845
<b>9</b>	-3,09	-3,09	-3,09	-1,06757049672431
<b>10</b>	-3,09	-3,09	-3,09	-3,09
<b>11</b>	-3,09	-3,09	-3,09	-0,565948784978835
<b>12</b>	-3,09	-3,09	-3,09	-1,06757049672431
<b>13</b>	-3,09	-3,09	-3,09	-0,180012358854845
<b>14</b>	-3,09	-3,09	-3,09	-0,565948784978835
<b>15</b>	-3,09	-3,09	-3,09	-1,06757049672431
<b>16</b>	-3,09	-3,09	-3,09	-0,180012358854845
<b>17</b>	-3,09	-3,09	-3,09	-0,565948784978835

## Anexo 7

Criterio	N-P	Promedio por aspectos
1	0,497520401580054	-2,58439262418108
2	1,00312777739898	-3,09
2	0,275630867112687	-2,36250308971371
4	0,372114973643685	-2,45898719624471
5	1,00312777739898	-3,09
6	1,00312777739898	-3,09
7	0,275630867112687	-2,36250308971371
8	0,275630867112687	-2,36250308971371
9	0,497520401580054	-2,58439262418108
10	1,00312777739898	-3,09
11	0,372114973643685	-2,45898719624471
12	0,497520401580054	-2,58439262418108
13	0,275630867112687	-2,36250308971371
14	0,372114973643685	-2,45898719624471
15	0,497520401580054	-2,58439262418108
16	0,275630867112687	-2,36250308971371
17	0,372114973643685	-2,45898719624471

Tabla de evaluación de los criterios.

MA- Muy Adecuado

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA

## Anexo 8

### Acta de aceptación del cliente.

ACTA DE ACEPTACION  
SISTEMA DE PRESERVACION DIGITAL

15 de junio de 2016

El centro de información científico-técnica, particularmente la Biblioteca de la UCI, tiene bajo su custodia parte importante de la producción científica de la UCI en formato digital. Debido al riesgo al que está expuesta esta información, ya sea por posible pérdida de la misma, o por la obsolescencia tecnológica a la que está expuesta, se dio inicio a un proyecto para la implantación de un sistema de preservación digital.

Luego de la revisión del sistema seleccionado, basado en el software libre Archivematica, se ha podido constatar que el mismo es compatible con los estándares de preservación digital mas importantes, tales como el modelo OAIS. Las pruebas funcionales realizadas al sistema han permitido comprobar que posee las funcionalidades adecuadas para realizar los procesos de preservación digital en la Biblioteca de la UCI. Luego de la implantación satisfactoria del mismo en la biblioteca, se procederá a su instalación en el nodo central de la universidad, de acuerdo a los procedimientos establecidos al efecto  
Sirva la presente como constancia de aceptación del sistema para su uso en la Biblioteca de la UCI.



Directora del CICT  
Laritza Martínez Negrin

