

**DISEÑO DE UN SISTEMA DE GESTION DOCUMENTAL Y PROTOTIPO DEL MODULO
DE MIGRACION PARA PRESERVACION DE LOS DOCUMENTOS**

SILVIA JULIANA ROSAS GUEVARA
JORGE ENRIQUE ESTRADA CORREA

Trabajo de grado para optar al título en Magíster en Tecnologías de la Información y la
Comunicación

Director
Sebastián Pérez Jiménez

UNIVERSIDAD PONTIFICIA BOLIVARIANA
ESCUELA DE INGENIERÍAS
MEDELLÍN
2018

NOTA DE ACEPTACIÓN

JURADO

JURADO

DIRECTOR

DEDICATORIA

Dedicamos este proyecto a nuestras familias, motor principal para cada día luchar por ser mejores personas y excelentes profesionales. A Dios y la Virgen María.

AGRADECIMIENTOS Y RECONOCIMIENTOS

Agradecimiento primeramente a Dios, nuestros familiares y todo el equipo de trabajo, sin ellos no hubiese sido posible la realización de este proyecto.

Reconocimiento especial a nuestro director de proyecto por su incondicional asesoría, dedicación y buenos aportes académicos.

Declaración de originalidad

Octubre 22 del 2018

Silvia Juliana Rosas Guevara

Jorge Enrique Estrada Correa

“Declaro que esta tesis (o trabajo de grado) no ha sido presentada para optar a un título, ya sea en igual forma o con variaciones, en esta o cualquier otra universidad” Art. 82 Régimen Discente de Formación Avanzada.

Firmas:



Two horizontal lines for signatures. The first line has a handwritten signature that appears to be 'Silvia'. The second line has a handwritten signature that appears to be 'Jorge'.

Índice General

Resumen	13
Glosario de Términos	14
Introducción.....	16
1. Planteamiento del problema, Objetivos y Justificación	18
1.1 Planteamiento del Problema	18
1.2 Objetivos.....	19
1.2.1 Objetivo General	19
1.2.2 Objetivos Específicos	19
1.3 Justificación	19
2. Antecedentes.....	21
2.1 Marco Teórico	21
2.1.1 Sistemas de Gestión y Administración Documental	21
2.1.1.1 Digitalización, Clasificación y Almacenamiento de los documentos	22
2.1.1.2 Administración de Expedientes Electrónicos y Flujo de Documentos	24
2.1.2 Modelo OAIS ISO 14721	26
2.1.3 PREMIS (Metadatos de preservación: Estrategias de ejecución)	29
2.2 Marco Legal	33
2.2.1 Archivo General de la Nación.....	33
2.2.2 Regulación Colombiana	35
2.2.3 Preservación del Patrimonio Digital, Criterios y Normas ISO	38
3. Metodología Aplicada.....	41
3.1 Metodología del Marco Lógico.....	41
3.1.1 Análisis	42
3.1.2 Diseño	42
3.1.3 Desarrollo.....	43
3.2 Metodología UML	43

3.3 Metodología para el desarrollo de software	43
3.4 Metodología de Minería de Datos	46
4. Requerimientos Funcionales del Sistema de Información Documental	49
5. Arquitectura, Modelo de Clases y Modelo de BD del Sistema de Información Documental	50
5.1 Requerimientos Arquitectónicos y Herramientas de Desarrollo	50
5.1.1 Diseño de Despliegue	50
5.1.2 Herramientas de Desarrollo	52
5.2 Diseño y Arquitectura del Sistema de Gestión Documental	53
5.2.1 Modelo Lógico.....	53
5.2.2 Modelo de Clases	53
5.2.3 Modelo de Base de Datos	54
6. Prototipo Funcional del Módulo de Migración del SGD.....	58
6.1 Arquitectura del Módulo de Migración.....	58
6.2 Diseño de los Formularios de la Aplicación	62
6.2.1 Layout.cshtml.....	62
6.2.2 Índice	63
6.2.3 Ingesta – Migración de Documentos	65
6.2.3.1 Clase PDF.....	67
6.3 Descripción Funcional	67
7. Pruebas del Prototipo Funcional de Migración	68
7.1 Plan de Pruebas	68
7.1.1 Preparar la Base de Datos	68
7.1.1.1 Series Documentales	68
7.1.1.2 Sub Series Documentales	68
7.1.1.3 Dependencias.....	69
7.1.1.4 Unidades Documentales	69
7.1.1.5 Listados de índices	70
7.1.1.6 Índices por Unidad Documental	70

7.1.1.7 Tablas de Retención Documental	71
7.1.1.8 Unidades Documentales por Tablas de Retención	71
7.1.2 Preparar los Documentos	72
7.1.3 Elaborar archivo plano con la información de los documentos a migrar	72
7.1.4 Preparar directorio destino	73
7.1.5 Configurar cadena de Conexión a la Base de Datos	74
7.1.6 Ejecutar Módulo de Migración	74
7.1.7 Analizar resultados obtenidos en el log de migración.	75
7.2 Análisis e interpretación de resultados.....	75
7.2.1 Resultados almacenados en las tablas de Auditoría	75
7.2.2 Proyecto PREMIS.Análisis	77
7.2.2.1 Listado de Documentos Migrados	78
7.2.2.2 Diseño de DashBoards	79
7.2.2.3 Consulta de Dashboards	84
7.3 Resumen de Evidencias y Resultados	85
8. Validación de Resultados con BI	86
8.1 Comprensión del problema	86
8.2 Comprensión de los datos	86
8.3 Preparación de los datos.....	89
8.3.1 Perfilamiento Tabla Resultados de Migración.....	89
8.3.2 Perfilamiento Tabla Valores Índices Migrados.....	90
8.4 Modelado	92
8.4.1 Modelo con Weka de los Resultados de la Migración.....	92
8.4.2 Modelo con Weka de los Valores Índices Migrados	96
8.5 Evaluación e Implementación	98
Conclusiones.....	99
Recomendaciones	101
Temas propuestos como continuación de este trabajo	101
Referencias bibliográficas	102

Anexos	106
---------------------	------------

Índice de Ilustraciones

Ilustración 1 – Expediente electrónico (MINTIC, 2017)	24
Ilustración 2 – Entidades funcionales de OAIS (Lavoie, 2014).....	29
Ilustración 3 – El modelo de datos PREMIS(Biblioteca Nacional de España, 2015).....	32
Ilustración 4 – Ciclo de Desarrollo Ágil (Trigas Gallego & Domingo Troncho, 2012)	44
Ilustración 5 – Niveles de abstracción de CRISP-DM (Gallardo, 2009)	47
Ilustración 6 – Modelo de proceso CRISP-DM (Gallardo, 2009).....	48
Ilustración 7 – Diagrama de Despliegue	52
Ilustración 8 – Proyectos de la solución.....	59
Ilustración 9 – Estructura del Proyecto PREMIS	60
Ilustración 10 – Diagrama Base de Datos Migración.....	61
Ilustración 11 – Modelo Entity Framework	62
Ilustración 12 – Plantilla principal (layout).....	63
Ilustración 13 – Formulario Inicial del Prototipo de Migración	63
Ilustración 14 – Ubicación archivo HomeController.cs	64
Ilustración 15 – Banners del Carrusel	64
Ilustración 16 – Opciones adicionales de la vista Índex	65
Ilustración 17 –Formulario para realizar el proceso de Migración	66
Ilustración 18 – Series Documentales	68
Ilustración 19 – Subseries Documentales.....	69
Ilustración 20 – Dependencias	69
Ilustración 21 – Unidades Documentales	69
Ilustración 22 – Listado de Índices	70
Ilustración 23 – Listado de Índices con su Unidad Documental	71
Ilustración 24 – Tabla de Retención Documental	71
Ilustración 25 – Unidades de la TRD Vigente	72
Ilustración 26 – Información de la TRD y valor de los Índices de cada documento.....	73
Ilustración 27 – Resultados de la Tabla LogMigracion	76
Ilustración 28 – Resultados de la Tabla Migracion_Layout.....	76
Ilustración 29 – Proyecto PREMIS.Analisis	77
Ilustración 30 – Formulario Principal del Proyecto PREMIS.Analisis	77
Ilustración 31 – Formulario listado de documentos migrados	78
Ilustración 32 – Vista de un documento seleccionado en la tabla.....	78
Ilustración 33 – Formulario de Diseño de Dashboards	79
Ilustración 34 –Nuevo Dashboard.....	80
Ilustración 35 –Selección origen de datos.....	80
Ilustración 36 –Asistente para la creación de consulta de datos	81
Ilustración 37 –Tablas para la creación de consultas al Dashboard	81
Ilustración 38 –Instrucción SQL generada.....	82

Ilustración 39 –Formulario para nuevo Dashboard	82
Ilustración 40 – Dashboard y sus propiedades	83
Ilustración 41 –Ejemplo Dashboard con una Rejilla.....	83
Ilustración 42 –Ejemplo Dashboard con una Grafica	84
Ilustración 43 –Consulta de un Dashboard	84
Ilustración 44 –Perfil DQ Analyzer Tabla Log Migración	89
Ilustración 45 –Perfil DQ Analyzer Tabla vIndicesDocumentoTexto	91
Ilustración 46 –Modelo con Weka de los Resultados de la Migración	93
Ilustración 47 –Técnica de Clasificación	94
Ilustración 48 –Árbol de decisión	95
Ilustración 49 –Técnica de Clúster.....	95
Ilustración 50 –Técnica de Clasificación	96
Ilustración 51 –Modelo con Weka de los Valores Índices Migrados.....	97
Ilustración 52 –Visualización de los atributos de los Valores Índices Migrados	97

Índice de Tablas

Tabla 1 – Campos de la tabla LogMigracion	87
Tabla 2 – Campos de la tabla vIndicesDocumentosTexto	88
Tabla 3 – Campos del archivo migración (metadatos).....	88

Resumen

Español

Con el avance de la tecnología, la digitalización de documentos ha incursionado en la sociedad de la información con el ánimo mejorar los procesos de administración documental en las organizaciones. El aumento masivo de la documentación electrónica y la disminución de los documentos en papel hizo necesario el desarrollo e implementación de nuevas innovaciones y herramientas tecnológicas para apoyar la administración y gestión documental. Actualmente se presentan grandes retos en los procesos de gestión, administración y conservación de los documentos digitales. Se hace necesario controlar los riesgos que afectan la seguridad, integridad y conservación de los documentos digitales a lo largo del tiempo.

Con este trabajo se aporta el diseño del sistema de gestión documental junto con el prototipo para el proceso de migración documental. Con él se ofrece una solución para garantizar la conservación y preservación de los documentos digitales, para aquellos que lo implementen como alternativa de gestión.

Inglés

With the advancement of technology, the digitalization of documents has entered into the information society with the aim of improving the processes of document management in organizations. The massive increase in electronic documentation and the reduction of paper documents made it necessary to develop and implement new innovations and technological tools to support document management and management. Currently there are major challenges in the processes of management, administration and preservation of digital documents. It is necessary to control the risks that affect the security, integrity and conservation of digital documents over time.

This work provides the design of the document management system together with the prototype for the documentary migration process. It offers a solution to ensure the preservation and preservation of digital documents, for those who implement it as an alternative management.

Glosario de Términos

AGN: Archivo general de la Nación. Es un establecimiento público del orden nacional encargado de formular, orientar y controlar la política archivística, coordinar el sistema nacional de archivos y la red nacional de archivos, y garantizar la conservación del patrimonio documental, asegurando los derechos de los ciudadanos y el acceso a la información, así como, el mejoramiento en la eficiencia de la gestión pública, la eficiencia de estado a través de una gestión documental articulada con el uso de las tecnologías de la información y las comunicaciones.

BPMN: Business Process Modeler Notation. Es un estándar para el modelado de procesos de negocio y servicios web; notación a través de la cual se expresan los procesos de negocio en un diagrama de procesos de negocio

CAD: Centro de Administración Documental.

CRISP-DM: Cross Industry Standard Process for Data Mining. Es la guía de referencia más ampliamente utilizada en el desarrollo de proyectos de Data Mining. Está dividida en 4 niveles de abstracción organizados de forma jerárquica en tareas que van desde el nivel más general, hasta los casos más específicos y organiza el desarrollo de un proyecto de Data Mining, en una serie de fases: comprensión del negocio o problema, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación.

MINTIC: Ministerio de Tecnologías de la Información y las Comunicaciones. Es la entidad que se encarga de diseñar, adoptar y promover las políticas, planes, programas y proyectos del sector de las tecnologías de la información y las comunicaciones. Dentro de sus funciones esta incrementar y facilitar el acceso de todos los habitantes del territorio nacional a las tecnologías de información y las comunicaciones y a sus beneficios.

OAIS: Open Archival Information System.

PDF: Portable Document Format, es un formato de almacenamiento para documentos digitales que independiente de las plataformas de software o hardware.

PREMIS: Preservation Metadata Maintenance Activity.

SCRUM: metodología de desarrollo ágil que tiene como base la idea de creación de ciclos breves para el desarrollo, comúnmente se llaman iteraciones (en SCRUM se llaman Sprints). Se puede dividir de forma general en 3 fases, que se entienden como reuniones, las cuales forman parte de los artefactos de la metodología junto con los roles y elementos que lo forman. Estas fases son: planificación del backlog, seguimiento del sprint y revisión del sprint.

SQL: Structured Query Language. Es un tipo de lenguaje vinculado con la gestión de bases de datos de carácter relacional, permite la especificación de distintas clases de operaciones entre éstas. El SQL brinda la posibilidad de realizar consultas con el objetivo de recuperar información de las bases de datos de manera sencilla.

TIFF: Tagged Image File Format. Formato de archivo informático para almacenar imágenes de mapa de bits.

TRD: Tablas de Retención Documental. Constituye un instrumento archivístico que permite la clasificación documental de la entidad acorde a su estructura orgánico-funcional, e indica los criterios de retención y disposición final resultante de la valoración documental por cada una de las agrupaciones documentales.

UML: lenguaje unificado de modelado. Es un estándar diseñado para visualizar, especificar, construir y documentar software orientado a objetos. Es esencial en la construcción del software.

Introducción

El soporte para el futuro y el progreso de la humanidad está dado por la conservación y preservación de los documentos en los diferentes sectores de la sociedad. En el sector empresarial, una compañía que pierda su historia documental tiene una alta probabilidad de estar destinada y condenada al cierre de sus operaciones; de igual forma, la humanidad sería relegada y dejaría de avanzar si perdiera su memoria por no contar con sus documentos y registros. De hecho, el modo de ver el universo podría ser hoy en día muy distinto, si muchos de los documentos que almacenaba la biblioteca de Alejandría no hubieran sido destruidos, este es uno de los hechos más nefastos para el progreso de la humanidad.(Consejo de la UE, 2002)

La invención del papel permitió el registro de las experiencias vividas y su conservación; con ello el conocimiento fue transmitido y expandido. Los constantes avances tecnológicos, herramientas e innovaciones dieron paso a la era digital, identificada con el documento digital. Este nuevo medio disponible para el registro de la información, brinda posibilidades no disponibles con el papel, entre otras: compartir al mismo tiempo un documento entre varias personas, realizar miles de copias de manera rápida y sin los costos económicos y ambientales, reducción de costos en la generación y almacenamiento, flujo electrónico de los documentos y el uso de firma digital por medio de certificados digitales. Los documentos digitales llevan consigo un alto riesgo para el acceso de la información y en el peor de los casos su pérdida por obsolescencia tecnológica o virus informáticos.

La mayoría del material bibliográfico y documental nace en un medio digital y no se conserva registro de dicha información en medios como el papel. El papel es un medio de conservación que podría ser mejor con miras a la preservación y acceso de la información, al no depender de ninguna tecnología, ni verse afectado por la obsolescencia tecnológica; pero actualmente solo se conserva los documentos en papel que fueron creados hace demasiado tiempo.

Hoy en día es una necesidad buscar cualquier recurso válido que permitan garantizar la preservación de los archivos digitales sin importar el continuo avance de innovaciones tecnológicas y la obsolescencia programada.

Este proyecto quiere aportar una alternativa de solución que dadas las condiciones en materia tecnológica, permita reducir el riesgo en la pérdida de la información almacenada en medios digitales. Para dar cubrimiento al desarrollo de la propuesta, el proyecto está planteado en 8 capítulos los cuales parten desde la definición del problema, hasta llegar a la interpretación de resultados y conclusiones. En el primer capítulo se identifica la problemática, se definen los objetivos, y se justifica el porqué del desarrollo de la propuesta. En el capítulo segundo, se aborda todo el estudio del marco teórico, normatividades y regulaciones, tanto como colombianas como a nivel mundial, para dar paso al capítulo 3 donde se analizan las metodologías relacionadas al diseño y desarrollo del SW, y el modelo de minería de datos que mejor se adapten a los requerimientos del proyecto.

El levantamiento de los requerimientos funcionales y arquitectónicos del sistema se establecen en el capítulo 4; el uso de la tecnología, la arquitectura y el diseño del modelo de la base de datos se contemplan en el capítulo 5. El desarrollo del prototipo funcional de migración se explica a nivel de detalle en el capítulo 6, y los resultados obtenidos de las pruebas realizadas al mismo se documentan en el capítulo 7; para finalizar con la validación con minería de datos en el capítulo 8.

1. Planteamiento del problema, Objetivos y Justificación

1.1 Planteamiento del Problema

Según (Natividad & Salamanca, 2015), “La conservación de los documentos digitales es un problema de gran importancia, así lo ha reconocido la UNESCO en la Carta para la preservación del patrimonio digital. En el capítulo 3 de la misma se reconoce el peligro de pérdida a que están sometidos estos materiales y se afirma: *“El patrimonio digital del mundo corre el peligro de perderse para la posteridad”*.”

Se hace necesario buscar alternativas, estrategias o el desarrollo de nuevas innovaciones, para mitigar y minimizar los riesgos ocasionados por la obsolescencia tecnológica, daños ocasionados por error humano, desastres naturales o ataques que aprovechan las vulnerabilidades en los sistemas de información (Cesar Tarazona, 2002). Alternativas que garanticen una alta probabilidad de preservar la documentación con el paso del tiempo. (Melorose, Perroy, & Careas, 2015)

Una de las mayores dificultades para preservar los documentos digitales tiene que ver con los procesos de migración documental, porque implican el cambio de sistemas o plataformas debido a la evolución de la tecnología; en este proceso es común encontrar problemas relacionados con el cifrado de los documentos, errores en los metadatos de los documentos o pérdida de información de los mismos, formato de archivos que debido a su obsolescencia ya no pueden ser leídos por los nuevos sistemas operativos o programas de software correspondientes, archivos corruptos que no se pueden leer, pérdida de documentos por daños, accidentes o delitos informáticos, delitos de adulteración de los documentos digitales y/o sus metadatos.

1.2 Objetivos

1.2.1 Objetivo General

Diseñar la arquitectura de un sistema de información para el almacenamiento y administración de documentos digitales con un prototipo funcional que cumpla con la normatividad vigente del sector publico colombiano.

1.2.2 Objetivos Específicos

- Identificar los requerimientos funcionales y arquitectónicos del sistema de información documental.
- Diseñar la arquitectura del sistema, modelo de clases y modelo de la base de datos del sistema de información documental.
- Implementar prototipo funcional del módulo de migración del sistema de información documental.
- Ejecutar plan de pruebas del prototipo funcional del módulo de migración del sistema de información documental.
- Validar resultados del plan de pruebas del prototipo funcional mediante el uso de herramientas de BI.
- Elaborar documentación del sistema de información documental.

1.3 Justificación

El activo más valioso para cualquier organización es su información, es muy difícil que una compañía pueda sobrevivir al nefasto evento de que su información histórica se pierda o no sea accesible por cualquier circunstancia; es un hecho que hoy dependamos cada vez más de los equipos de cómputo para convertir los documentos físicos en digitales. Es de vital importancia para cualquier organización contar con políticas de seguridad, implementar sistemas de información y establecer controles que mitiguen el riesgo a la pérdida de la

información, mediante medidas que garanticen la continuidad del negocio ante la materialización de cualquier tipo de riesgos asociados al daño, adulteración o pérdida de los documentos en general.

Al momento de implementar una herramienta de software, las empresas se ven obligadas a buscar alternativas entre los productos existentes en el mercado o llevar a cabo todo el proceso de análisis, diseño y programación de un producto hecho a la medida de sus necesidades. Si bien, en la actualidad existen soluciones para la administración documental, ninguna de éstas contempla en sus productos las herramientas necesarias para que sean configurable desde el momento de la implementación y que garanticen los procesos de migración de documentos cuando se requiera de un cambio en la plataforma tecnológica.

Debido a los constantes cambios en las herramientas de cómputo tanto en hardware como en software por la obsolescencia programada, la aparición de nuevas innovaciones tecnológicas, cambios de paradigmas y otras circunstancias; los “Almacenes o bodegas de documentos” administrados por algún software e implementados sobre una determinada plataforma tecnológica, deben ser migrados periódicamente a otras plataformas.

Este proceso de migración suele ser muy costoso en recurso, dinero y pone en riesgo a los mismos documentos (daños ocasionados por procesos de descifrado y cifrado o cambios en los parámetros).

Entre las grandes dificultades a las que se ven enfrentados los ingenieros y tecnólogos a la hora de emprender los procesos de migración, tienen que ver con:

- Descifrado de los documentos: Algunos sistemas de información cifran los documentos utilizando para ello un algoritmo de cifrado propio o abierto, pero sólo el fabricante del software conoce el algoritmo requerido para descifrar la clave y semilla.
- Estructura de Datos (Modelo Entidad Relación) de la base de datos del sistema obsoleto: Se es necesario conocerla y realizar allí las consultas para obtener los índices, las tablas de retención documental y los metadatos que se tengan para cada uno de los documentos a ser migrados.

El proceso de migración documental se debe realizar de manera transparente, eficaz y segura, este proceso no debería depender del proveedor del sistema en el que se crearon los documentos digitales. De hecho, al momento de realizar el proceso de migración, es necesario conocer la arquitectura de la herramienta en la que se crearon y almacenaron los documentos que se desean migrar (modelo de datos, funciones, algoritmos de cifrado, parámetros del sistema, mecanismos de seguridad) estos elementos por lo general sólo lo conocen las empresas que desarrollaron el sistema de información. Es importante analizar el escenario sobre un cambio sobre los sistemas cuando el proveedor ya no está disponible y a qué riesgos se enfrentan las entidades que dependen de terceros.

Para brindar una alternativa de solución a la problemática de la gestión documental y sus procesos de migración, se diseña la arquitectura de un sistema de información para la gestión de documentos digitales. Este diseño contempla los niveles mínimos de seguridad para garantizar la integridad, conservación de la información, y permite que los procesos de migración sean confiables y eficaces. El modelo está ajustado a los estándares de calidad ISO relacionados con la administración y preservación documental y cumple con la normatividad vigente para el sector público colombiano (Archivo General de la Nación).

2. Antecedentes

2.1 Marco Teórico

Las siguientes referencias, definiciones y modelos permiten dar contexto de los sistemas de gestión documental y las buenas prácticas para la conservación del patrimonio digital.

2.1.1 Sistemas de Gestión y Administración Documental

Los sistemas para la gestión y administración documental (Rodríguez A & Gaitán Gómez, 2011) son herramientas de cómputo destinadas a tareas de:

- Digitalización, clasificación, almacenamiento de los documentos.
- Administración de expedientes electrónicos.
- Búsquedas y consultas de los expedientes y documentos.

- Registro de préstamos de expedientes y/o documentos.
- Flujo electrónico de documentos.

El documento electrónico se define como *“La información generada, enviada, recibida, almacenada y comunicada por medios electrónicos, ópticos o similares”* (MINTIC, 2017).

2.1.1.1 Digitalización, Clasificación y Almacenamiento de los documentos

La digitalización de los documentos es un proceso que permite generar un documento en formato digital a partir de un documento que se encuentra en un soporte físico por lo general en papel. Este proceso requiere de actividades previas que permitan que el resultado final sea un documento de calidad, con unas condiciones mínimas que permitan su lectura adecuadamente desde cualquier dispositivo electrónico. Generalmente, el Centro de Administración Documental (CAD) es el área encargada de realizar estas actividades, las cuales inician con acondicionamiento del documento original en papel que le permita pasar por el scanner sin ningún inconveniente, luego de alistar y radicar el documento, este es procesado por el scanner y genera como resultado un archivo en algún formato (TIFF, PDF, entre otros), actualmente, el formato más recomendado es el archivo PDF/A (Archivo General de la Nación, 2012).

No todos los documentos digitales tienen su origen a partir de un documento en papel que son digitalizados, hoy por hoy, muchos documentos son creados desde programas o herramientas de software y su origen está atado digitalmente a un archivo electrónico, ejemplos de esto, son los correos electrónicos, documentos generados en formato PDF los cuales nunca se imprimen y que a la vez son radicados, firmados por medio de certificado digitales y almacenamos en servidores de archivos o en bases de datos.

Los sistemas de gestión y administración documental por lo general proporcionan opciones y herramientas que permiten diseñar plantillas de documentos a partir de las cuales los usuarios pueden generar documentos y manejar todo el ciclo de vida de estos (creación, revisión, aprobación y publicación), igualmente, permiten los procesos de radicación, clasificación y almacenamiento de los metadatos respectivos y el firmado digital.

Los documentos deben ser clasificados de acuerdo con la definición dada por el CAD en las tablas de retención documental. (Archivo General de la Nación, 2001). Las tablas de retención documental permiten agrupar los documentos en una estructura jerárquica de tres niveles, denominados: series documentales, las cuales a su vez se clasifican en subseries documentales y ellas se clasifican en unidades documentales y/o tipos documentales. Las tablas de retención y valoración documental deben ser establecidas como se mencionó con anterioridad por el CAD y deben ser actualizadas periódicamente debido a cambios normativos o por necesidades al interior de las entidades. En estas tablas de retención se establece también para cada unidad documental los tiempos asociados a las diferentes etapas por las que pasan los documentos.

Las etapas que componen el ciclo de vida de los documentos establecidas en las tablas de retención documental se pueden clasificar como: archivo de gestión, archivo central, archivo histórico.

(Edin Gutiérrez, 2017)

Por otro lado, a cada documento se le deben asignar una serie de datos conocidos como índices, los cuales se configuran para cada unidad documental, éstos índices son los que permiten que, por medio de ellos, un usuario pueda luego buscar los documentos.

Una vez que el documento se encuentra en formato digital, se debe almacenar al igual que sus índices de búsqueda en algún tipo de repositorio. Los índices son almacenados en algún motor de base de datos relacional o de cualquier otro tipo, y los documentos se pueden almacenar en disco duro u otros medios como: servidor de archivos, discos ópticos, base de datos, en la nube.

Los documentos que nacen digitalmente deben tener la posibilidad de recibir una firma certificada por medio de certificados digitales expedidos por una entidad certificadora, de no ser así, el usuario tendría que imprimir el documento, firmar el documento una vez impreso, digitalizarlo con la firma manuscrita y volverlo a subir a la aplicación; lo anterior involucra muchos reproceso y gastos innecesarios.

2.1.1.2 Administración de Expedientes Electrónicos y Flujo de Documentos

Un expediente se puede definir como una colección de documentos relacionados con un tema o trámite en particular y que se conservan manteniendo la integridad y orden en que fueron producidos desde el inicio hasta el final del trámite. Se debe considerar, que en ocasiones, un mismo documento puede hacer parte de varios expedientes. Cada entidad o empresa debe administrar los expedientes de acuerdo con las directrices y normas propias de cada país. (MINTIC, 2017)

En Colombia, el Artículo 6° del Acuerdo 003 de 2015 expedido por el Archivo General de la Nación define el expediente electrónico como el *“Conjunto de documentos y actuaciones electrónicas producidos y recibidos durante el desarrollo de un mismo trámite o procedimiento, acumulados por cualquier causa legal, interrelacionados y vinculados entre sí, manteniendo la integridad y orden dado durante el desarrollo del asunto que les dio origen y que se conservan electrónicamente durante todo su ciclo de vida, con el fin de garantizar su consulta en el tiempo”*.

El expediente debe contener cómo mínimo los siguientes elementos (Ver Ilustración 1): documentos electrónicos de archivo que hacen parte del expediente, foliado electrónico, índice electrónico, firma del índice, electrónico y metadatos.

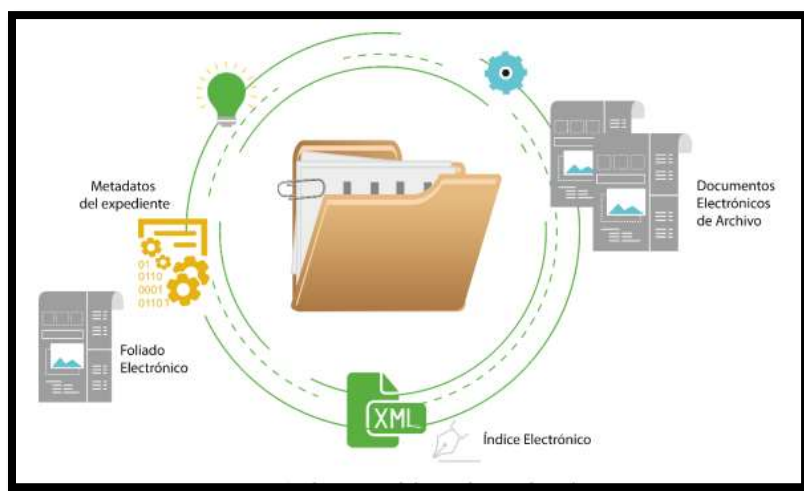


Ilustración 1 – Expediente electrónico (MINTIC, 2017)

El foliado electrónico consiste en la asociación de todos los documentos que hacen parte del expediente a un índice electrónico el cual deberá ser firmado digitalmente por la autoridad cuando el expediente sea cerrado. Lo anterior con el fin de garantizar la integridad y seguridad del expediente electrónico. El índice electrónico, *“Constituye un objeto digital que contiene la identificación de los documentos electrónicos que componen el expediente, debidamente ordenados en orden cronológico, para reflejar la disposición de los documentos, así como otros datos con el fin de preservar la integridad y permitir la recuperación del mismo”* (MINTIC, 2017).

El índice electrónico se deberá firmar digitalmente al cierre del expediente, sin perjuicio de las garantías de seguridad de la información que deberán adoptar las entidades públicas durante la etapa de tramitación. En concordancia con el Parágrafo del art. 23 del Acuerdo 02 de 2014 del AGN.

Tanto los documentos electrónicos de archivo como sus expedientes necesitan la asignación de metadatos mínimos obligatorios siguiendo la norma ISO 23081, DUBLIN CORE (SEDIC, 2003) y METS.

Deben existir mecanismos o herramientas para que un usuario que posea los permisos necesarios pueda realizar consultas en línea de documentos y expedientes electrónicos. Las búsquedas se deben realizar utilizando formularios que permitan ingresar los criterios de búsqueda deseados y la aplicación con estos criterios debe ir a buscar los documentos que en sus índices de búsqueda cumplan con ellos.

Como un valor agregado, estos sistemas de información pueden incorporar la funcionalidad que le permita al CAD registrar y llevar el control de los expedientes físicos y/o documentos físicos a usuario que por algún motivo requieran consular los documentos que se encuentran almacenados también en soporte físico.

Es muy importante que estos sistemas de información cuenten también con el mecanismo que permita que los documentos que se encuentran en la etapa de gestión en el ciclo de vida documental puedan fluir de manera controlada por medio de flujos de trabajo (WorkFlow) preconfigurados de acuerdo con tipos de trámite.

Se utilizan herramientas de tipo BPMN, Modelo y Notación de Procesos de Negocio, es un estándar para el diseño gráfico que permite el modelado de procesos de negocio, en un formato de flujo de trabajo (WorkFlow). Una vez modelados los procesos, se pueden iniciar trámites dentro de la entidad que fluyen por dichos procesos y cada trámite va acompañado de los documentos digitales que lo soporta.

2.1.2 Modelo OAIS ISO 14721

OAIS ISO 14721(ISO, 2015) es el modelo de referencia que define los procesos necesarios para preservar y acceder a los objetos de información de forma efectiva y a largo plazo, y establece un lenguaje común que los describe”(Cruz Mundet & Díez Carrera, 2016), y aclara el autor los dos siguientes aspectos:

- *“Se trata de un modelo, como su propio nombre indica, es decir, no especifica una forma de hacerlo; sino que proporciona el marco para poder hacerlo con éxito, describiendo las funcionalidades básicas y los tipos de información requeridos para el entorno de preservación.”*
- *“El término abierto se usa para dar a entender que se ha desarrollado en foros abiertos y no implica que el acceso al archivo sea ilimitado”.*

La National Aeronautics and Space Administration (NASA), desarrolló este modelo con las características que debe tener un sistema de preservación de documentos digitales a largo plazo, a raíz de la necesidad de conservar con seguridad sus documentos digitales, este modelo fue llamado *“Reference Model for an Open Archival Information System (OAIS)”*. En el 2003, OIAS se transformó en una Norma ISO con el código 14721:2003 y en junio de 2012, se realiza una nueva revisión y actualización UNE-ISO 14721:2015.

El modelo OAIS (Lavoie, 2014) ha sido ampliamente aceptado y utilizado en el ámbito de los CAD (Centros de Administración Documental) y es el modelo que propone el Ministerio de Tecnologías de la Información y las Comunicaciones MinTic para la implementación por parte de las entidades públicas, lo anterior puede consultarse en la Guía No. 6 - Sistemas de Gestión de Documentos Electrónicos: *“En lo referente a requisitos de la gestión de los*

documentos, aspecto en el que este documento no profundiza, se recomienda seguir el modelo de la norma ISO 14721, basado en el modelo del Open Archive Information System –OAIS”. (MINTIC, 2011).

Se busca por medio de la implementación de este modelo, que los documentos almacenados cumplan con los principios básicos de la seguridad de la información: integridad, disponibilidad, autenticidad, niveles de accesibilidad y que conserve estas características de seguridad a lo largo del tiempo. La integridad garantiza que el documento nunca haya sido alterado o adulterado, la disponibilidad garantiza que el documento siempre estará disponible para su uso y/o consulta, un documento es auténtico cuando existe certeza sobre la persona que lo ha elaborado, manuscrito o firmado, los niveles de permisos y accesibilidad permiten que los documentos sólo puedan ser accedidos por las personas que poseen los privilegios adecuados.

Este modelo se basa en seis procesos o funcionalidades (Ver Ilustración 2):

1. Adquisición (Ingesta), donde recibe la información de los productores, la prepara para almacenarla y gestionarla en el repositorio.
2. Gestión de archivo (Archival Storage), que define el uso del repositorio, el almacenamiento de los documentos, el mantenimiento de paquetes de información de archivo, responsable de asegurar que el contenido que se ha archivado se ha hecho de forma adecuada para garantizar su mantenimiento a largo plazo.
3. Gestión de los datos, que permite describirlos. Proporciona los servicios y las funciones para cumplimentar, mantener y acceder a información descriptiva (Descriptive Information), que idéntica y documenta los fondos de archivo, así como los datos administrativos usados para gestionar el archivo. Incluye administrar la base de datos y mantenerla actualizada, recibir y responder consultas, y generar informes.
4. Administración que gestiona las operaciones diarias del archivo y vela por la seguridad de los documentos. Se ocupa de la operativa global del OAIS. Comprende negociar los acuerdos de transferencia, gestionar la configuración del sistema, actualizar la información del archivo, controlar el acceso físico, establecer normas y políticas, auditar las transferencias, activar las solicitudes y mantener el servicio al cliente.

5. Acceso que ayuda a los usuarios a identificar y obtener la información y permite su consulta.
6. Plan de Preservación, cada entidad debe elaborar y ejecutar un plan de preservación a largo plazo.

El objetivo del OAIS: la preservación a largo plazo, se tropieza con una serie de factores negativos como la naturaleza cambiante de la industria informática, la naturaleza efímera de los medios de almacenamiento de datos electrónicos y la del propio software, así como los cambios en la base de conocimiento de la comunidad designada. Entre las posibilidades existentes, la estrategia propuesta se basa en la migración, porque mejora la rentabilidad, facilita adaptarse a las nuevas tecnologías y evita el deterioro de los medios. Con todo esto, las migraciones llevan tiempo, costos y exponen al OAIS a una mayor probabilidad de perder información. El modelo analiza los distintos tipos de migración (refresco, replicado, reempaquetado y transformación) y valora sus pros y sus contras, sin optar por ninguno de ellos en concreto. Asimismo, incluye entre los objetivos de la preservación los servicios de acceso y uso, dentro de los que reconoce la importancia de respetar la familiaridad de los usuarios con las interfaces a través de las que operan, algo que muchas veces no se tiene en cuenta.

Para conseguir dar cumplimiento a estas funcionalidades, el modelo define tres etapas diferenciadas: la adquisición, la conservación y el acceso. (Raventós, 2010)

La conservación se puede conseguir gracias al repositorio de archivo y a la migración hecha en la etapa anterior. Una vez depositados en el repositorio de manera periódica hay que ejecutar acciones de detección de errores y de garantía de legibilidad. Se trata de conseguir que el sistema de archivo a largo plazo no quede obsoleto tecnológicamente, y que conserve los objetos digitales estables, igual que los mecanismos que permiten su uso. Waters y Garrett (Group, 1996) proponen la siguiente definición de migración *“Un conjunto de tareas diseñadas para alcanzar la transferencia periódica de materiales digitales de una configuración hardware/software a otra, o desde una generación tecnológica de ordenador a una generación subsiguiente”*.

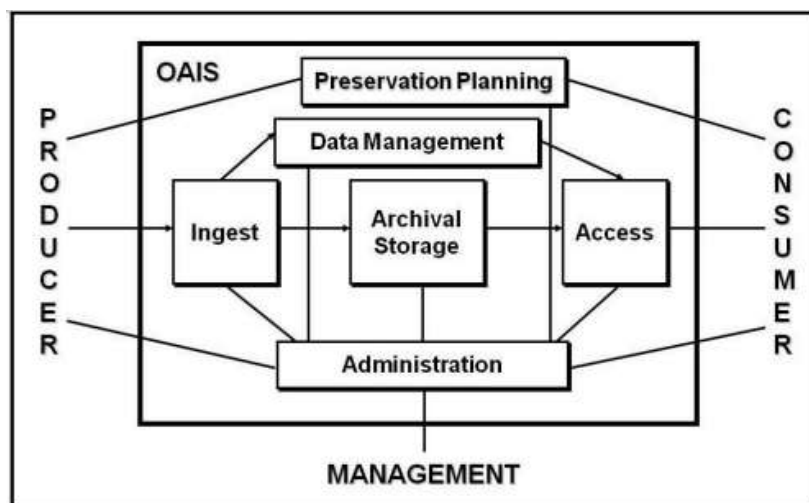


Ilustración 2 – Entidades funcionales de OAIS (Lavoie, 2014)

2.1.3 PREMIS (Metadatos de preservación: Estrategias de ejecución)

Dentro del Plan de Preservación propuesto por OAIS se deben incorporar los mecanismos, herramientas y procedimientos que se deben adoptar para realizar los procesos de migración documental, por otra parte, la Norma ISO 18492(ICONTEC, 2013) establece: *"Usar metadatos para la conservación de la información. Para que la información digital siga siendo accesible, se deben abordar algunos metadatos técnicos ante los cambios tecnológicos tanto estructurales y de almacenamiento como de la gestión. Estos metadatos deben abordar el acceso, la seguridad, la migración, la conservación y las transferencias, para asegurar la autenticidad, fiabilidad, integridad y disponibilidad de la información. Para ello, se aconseja el uso de metadatos en el momento de la incorporación, como la fecha y hora de creación del documento, la identificación de los agentes implicados en su creación, y la documentación o registro de su estructura, forma, propiedades físicas, dependencias técnicas, relación con otros documentos y requisitos para su reproducción"*, también recomienda asegurar que los metadatos sean interoperables. En las migraciones y los cambios de soporte o de formato, las aplicaciones de software deben permitir la creación automática de metadatos, como el tamaño del archivo, su formato, sus datos, la firma digital u otros atributos similares de autenticidad del documento. Además, la aplicación del software debería permitir la entrada manual de otros metadatos, como la clasificación, el periodo de

conservación, la serie de los documentos y las palabras clave, los cuales pueden ser necesarios, especialmente cuando estos documentos electrónicos se quieran recuperar. Para ello, los depósitos de almacenamiento deben garantizar que las herramientas de apoyo a la captura y el uso de metadatos sean suficientemente flexibles y escalables, a fin de poder acomodar elementos de metadatos más ricos a medida que estén disponibles. (Giménez Chornet, 2014)

Los metadatos (Agudelo, 2009) como lo define el autor: “*consiste en un conjunto de atributos o elementos necesarios para describir un recurso determinado, que funciona como identificador de los materiales digitales diseñados*”.

Los metadatos son utilizados con el fin de:

- Recuperar la información.
- Administrar los documentos.
- Gestión de derechos, autoría y propiedad intelectual.
- Control y descripción de procesos.
- Seguridad y autenticación.
- Preservación y conservación.

Los metadatos se pueden clasificar de acuerdo a su finalidad:

- Administrativos: Utilizados para:
 - Adquisición de información
 - Control de versionamiento
 - Requerimientos legales para el acceso
 - Derechos y reproducción.
- Descriptivos: Se utilizan con el fin de:
 - Proporcionar ayuda en la búsqueda
 - Índices especializados
 - Anotaciones de los usuarios
- Técnicos: Se usan para:
 - Documentación de hardware y software

- Digitalización de la información
- Autenticación y datos de seguridad
- Control de tiempo de respuesta de sistemas
- De Preservación: Son útiles para:
 - Informar sobre las condiciones de los recursos físicos
 - Informar sobre las acciones llevadas a cabo para preservar versiones físicas y digitales de recursos
- Uso:
 - Información sobre versiones
 - Reutilización del contenido del recurso

Para el manejo de metadatos la comunidad ha trabajado en diferentes frentes donde han surgido diferentes alternativas y estándares, según la norma UNE-ISO 23081-1: 2008, los metadatos son *"Información estructurada o semi estructurada que posibilita la creación, registro, clasificación, acceso, conservación y disposición de los documentos a lo largo del tiempo"* (Archivo General de la Nación, 2014)

"Los metadatos de preservación soportan las actividades cuyo objetivo es asegurar la utilización a largo plazo de un recurso digital" (Giusti, 2015).

- **Dublin Core:** Modelo de metadatos elaborado por la DCMI (Dublin Core Metadata Initiative) abocado a definir el conjunto de elementos básicos para describir los recursos electrónicos y facilitar su recuperación.
- **METS:** Metadata Encoding and Transmission Standard (METS). Esquema para describir objetos de bibliotecas digitales complejas que utiliza el lenguaje XML y asocia metadatos administrativos y descriptivos.

PREMIS fue elaborado por un grupo de trabajo internacional patrocinado por OCLC y RLG desde 2003-2005. Este grupo de trabajo elaboró un informe denominado PREMIS Data Dictionary for Preservation Metadata (Diccionario de datos PREMIS de metadatos de preservación), este grupo conformado por participantes de diferentes ámbitos para la utilización de metadatos aplicados en actividades de preservación digital. En mayo del 2005 publica el informe: diccionario de datos para metadatos de preservación. (Caplan, 2009)

Los metadatos de preservación que define el Diccionario de Datos presentan las siguientes características (Biblioteca Nacional de España, 2015)

- Contribuyen a la viabilidad, disponibilidad, claridad, autenticidad y la identidad de los objetos digitales en un contexto de preservación.
- Representan la información que la mayoría de los repositorios necesitan conocer para preservar materiales digitales a largo plazo.
- Prestan especial atención a los «metadatos aplicables»: rigurosamente definidos, basados en directrices para su creación, gestión y uso, y orientados a flujos de trabajo automatizados.
- Son técnicamente neutros, es decir, no se presupone el uso de ninguna tecnología de preservación en concreto, ni tampoco de estrategias, sistemas de almacenamiento y gestión de metadatos, etc.

El diccionario PREMIS define un conjunto de unidades semánticas, propiedades, e información que la mayoría de los repositorios necesita conocer de sus entidades para asegurar su preservación.

Se desarrolló un modelo de datos para organizar las unidades semánticas definidas en el diccionario de datos. El modelo define cinco entidades (Ver Ilustración 3): intelectuales, objetos, acontecimientos, derechos y agentes.

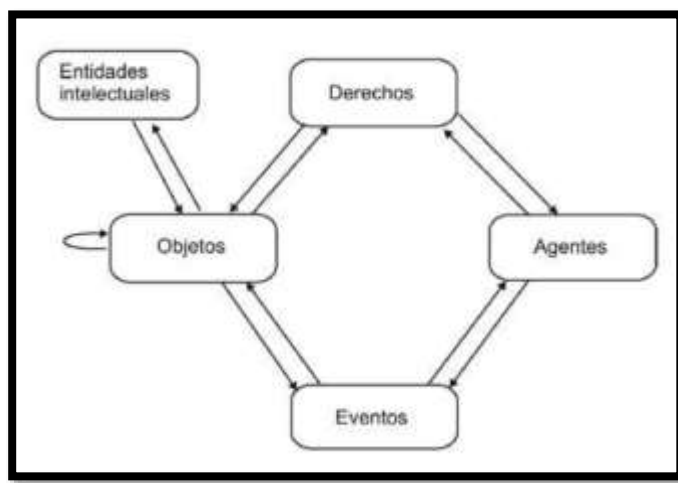


Ilustración 3 – El modelo de datos PREMIS(Biblioteca Nacional de España, 2015)

- **Entidad Intelectual:** conjunto de contenidos que se considera una única unidad intelectual a efectos de gestión y descripción. Una Entidad Intelectual puede comprender otras Entidades Intelectuales.
- **Objeto:** unidad discreta de información en formato digital
- **Acontecimiento:** acción que al menos afecta a un Objeto o Agente asociado o conocido por el repositorio de preservación.
- **Agente:** persona, organización o programa/sistema informático asociado a los Acontecimientos durante la vida de un Objeto, o a los Derechos ligados a un objeto.
- **Derechos:** declaración de uno o varios derechos o permisos pertenecientes a un Objeto o Agente.

PREMIS se utiliza fundamentalmente para el diseño de los repositorios, para su evaluación y para el intercambio de los paquetes de información archivada entre los repositorios de preservación.

Los diseñadores y/o desarrolladores de aplicaciones de software de repositorios de preservación deben utilizar PREMIS como una guía sobre la información que dichas aplicaciones deben obtener y registrar o que de alguna manera debe tenerse en cuenta para la gestión del repositorio.

Quienes tengan previsto implementar un repositorio de preservación, deben utilizar PREMIS como una lista de control para evaluar los posibles softwares. Los sistemas que soporten el Diccionario de datos de PREMIS serán más capaces de preservar los recursos de información a largo plazo. (Caplan, 2009)

2.2 Marco Legal

2.2.1 Archivo General de la Nación

En Colombia, el Archivo General de la Nación (AGN) es la entidad pública del orden nacional responsable de formular, orientar y controlar la política archivística, coordinar el

sistema nacional de archivos y garantizar la conservación del patrimonio documental; asegurando los derechos de los ciudadanos y el acceso a la información. (Archivo General de la Nación, 2017)

Bajo las directrices dadas por el AGN, el proceso archivístico está compuesto de las siguientes actividades o subprocesos:

- Producción o Recepción (Radicación de los documentos)
 - Distribución
 - Trámites
 - Consulta
 - Organización
 - Disposición Final
 - Conservación y preservación del Documento.
- a) Producción o Recepción: En esta etapa, se lleva a cabo todos los pasos necesarios para recibir bien sea en formato físico o digital el documento, verificar formato y validez de acuerdo a las políticas de la entidad, realizar el proceso de radicación del documento por algún medio válido, registrar el radicado generado, preparar y digitalizar el documento si éste viene en formato físico (papel), registrar los índices de búsqueda o metadatos y almacenar el documento cifrado y con todos los requerimientos de seguridad bien sea en un servidor de archivo o en la mismos gestor de base de datos del sistema.
- b) Distribución: Procedimiento por el cual un documento que ingresa o se genera en la entidad es entregado físicamente o direccionado por algún mecanismo o herramienta de software a la persona requerida.
- c) Trámites: Cuando él envío de un documento (distribución) se realiza por medio de sistema Workflow (BPMN u otro). Es normal que el software de gestión documental posea un módulo para la configuración y administración de flujos de trabajo o procesos. Mediante estas herramientas los documentos que son digitalizados realizan todo el

proceso de gestión de los trámites o expedientes electrónicos al que están asociados de manera virtual, fluyendo el documento por todas las dependencias y/o funcionarios de acuerdo con el flujo o proceso preestablecido, evitando con esto el flujo mismo del documento físico (cuando es este su formato original).

- d) Consulta: Establece los lineamientos o mecanismos para que los usuarios y/o funcionarios de la entidad puedan tener acceso a los documentos y realizar la consulta de los mismos o de los expedientes asociados con dichos documentos. Cuando la entidad posee un sistema de gestión documental lo normal es que este sistema posea las interfaces necesarias para que los usuarios al interior o al exterior puedan realizar la búsqueda de expedientes o de documentos, la búsqueda se realiza por medio de los metadatos y de acuerdo con los privilegios o permisos que dicho usuario posea sobre dicho tipo u unidad documental.
- e) Organización: Se realiza la clasificación documental de acuerdo con lo establecido en las Tablas de Retención y Valoración Documental, este procedimiento se realiza por lo general en la etapa de digitalización e indexación del documento cuando la misma se realiza a través de un sistema de gestión documental. Cuando un documento hace parte de un expediente, se realiza además el proceso de numeración o foliado de dicho documento dentro del expediente.
- f) Conservación (Preservación) y Disposición final: Se realizan procesos de reproducción, duplicación y migración de los documentos. En este proyecto uno de los objetivos está centrado precisamente, en el proceso de Migración.

2.2.2 Regulación Colombiana

Las siguientes normas permiten la regulación la gestión de los archivos y los sistemas de gestión documental:

Ley 80 de 1989; establece las funciones principales del Archivo General de la Nación. (Congreso de la República de Colombia., 1989)

Ley 223 de 1995; se centra en los principios de la gestión de documentos y establece los principios básicos para que las organizaciones puedan establecer un marco de buenas prácticas que mejore de forma sistemática y efectiva la creación y mantenimiento de sus documentos, apoyando la política y los objetivos de la organización. (Alonso, García, & Lloveras, 2007)

Decreto 2150 de 1995; por el cual se suprimen y reforman regulaciones, procedimientos o trámites innecesarios existentes en la Administración Pública. Estableció que las entidades públicas deberán habilitar sistemas de transmisión electrónica de datos para que los usuarios envíen o reciban información requerida en sus actuaciones frente a la administración. (Presidente de la República de Colombia, 1995)

Circular 02 de 1997, establece parámetros para tener en cuenta en la implementación de nuevas tecnologías en los archivos públicos. (Colombia, 1997)

Ley 527 de 1999, por medio de la cual se define y reglamenta el acceso y uso de los mensajes de datos, del comercio electrónico y de las firmas digitales, y se establecen las entidades de certificación y se dictan otras disposiciones. Esta ley hace referencia a ciertos documentos que deben ser conservados, siempre y cuando cumplan las condiciones necesarias; por ejemplo, el mensaje de datos o el documento debe ser conservado en el formato que se haya generado, enviado o recibido o en algún formato que permita demostrar que reproduce con exactitud la información generada, enviada o recibida. En su artículo 2, que define los conceptos de: mensaje de datos, comercio electrónico, firma digital, entidad de certificación, intercambio electrónico de datos y sistema de información. En su artículo 5, establece “Reconocimiento jurídico de los mensajes de datos. No se negarán efectos jurídicos, validez o fuerza obligatoria a todo tipo de información por la sola razón de que este en forma de mensaje de datos.” (Congreso de la República de Colombia., 1999)

Ley 594 de 2000, por medio de la cual se dicta la Ley General de Archivos y se dictan otras disposiciones. SE encuentra reglamentada parcialmente por el decreto 4124 de 2004. En su artículo 19, establece “Soporte Documental. Las entidades del estado podrán incorporar

tecnología de avanzada en la administración y conservación de sus archivos, empleando cualquier medio técnico, electrónico, informático, óptico o telemático, siempre y cuando cumplan con los requerimientos contemplados en la ley. (Congreso de la República de Colombia., 2000)

Decreto 1747 de 2000, por el cual se reglamenta parcialmente la ley 527 d 1999, en lo relacionado con las entidades de certificación, los certificados y las firmas digitales. Este decreto define conceptos como repositorio en los sistemas de información, certificado, estampado cronológico y entidad de certificación. (Presidente de la República de Colombia, 2000)

Acuerdo 060 de 2001, por el cual se establecen pautas para la administración de las comunicaciones oficiales en las entidades públicas y las privadas que cumplen funciones públicas. Define el concepto de documento electrónico de archivo. (Colombia, 2001)

Acuerdo 027 de 2006, por el cual se modifica el acuerdo No. 7 del 29 de junio de 1994: Modifica el glosario reglamento general de archivos. Definiciones de documento electrónico de archivo y archivo electrónico. (Colombia, 2006)

Ley 1437 de 2011, por la cual se expide el código de procedimiento administrativo y de lo contencioso Administrativo: Introduce los conceptos de expediente electrónico, sede electrónica, notificación electrónica, entre otras. (Congreso de la República de Colombia., 2011)

Ley 1564 de 2012, por medio de cual se expide el Código General del Proceso y se dictan otras disposiciones. Dicha ley introduce elementos como el manejo de expedientes y uso de las tecnologías de la información y las comunicaciones en todas las actuaciones judiciales. El artículo 74 determina que se podrá conferir poder especial por mensaje de datos con firma digital.

Decreto 2693 de 2012, por el cual se establecen los lineamientos generales de la Estrategia de Gobierno en Línea de la República de Colombia, se reglamentan parcialmente las Leyes 1341 de 2009 y 1450 de 2011. (Congreso de la República de Colombia., 2012b)

Circular 002 de 2012, “Adquisición de Herramientas Tecnológicas de Gestión Documental”. Hace referencia a la circular No. 004-2010 de Coinfo. “Estándares mínimos en procesos de administración de archivos y gestión de documentos electrónicos.” En particular, la recomendación No. 4 establece: Con el fin de preservar en el tiempo los documentos digitales producidos a recibidos por las diferentes entidades del Estado, y en particular aquellos que conforman el patrimonio documental del País, es necesario que se adopten estándares apropiados que aseguren la conservación a largo plazo de los documentos electrónicos. (Colombia, 2012b)

Acuerdo 004 de 2013, por el cual se reglamenta parcialmente los decretos 2578 y 2609 de 2012 y se modifica el procedimiento para la elaboración, presentación, evaluación, aprobación e implementación de las Tablas de Retención Documental y las Tablas de valoración Documental. Art 14. Actualización, Cuando se transformen tipos documentales físicos en electrónicos. Art. 18. Usos de tecnologías de automatización de la elaboración, consulta y actualización de la TRD, de forma que facilite la interoperabilidad con el registro único de Series Documentales y otros sistemas de información de cada entidad. (Colombia, 2012a)

Acuerdo 005 de 2013, por el cual se establecen los criterios básicos para la clasificación, ordenación y descripción de los archivos en las entidades públicas y privadas que cumplen funciones públicas y se dictan otras disposiciones. Art 20. Utilización de medios tecnológicos en la descripción archivística. Art 21. Compactibilidad e interoperabilidad. Art 22. Transferencia de Información. (Colombia, 2013)

2.2.3 Preservación del Patrimonio Digital, Criterios y Normas ISO

Se definió a la preservación digital como el conjunto de prácticas de naturaleza política, estratégica y acciones concretas, destinadas a asegurar el acceso a los objetos digitales a largo plazo. (Giusti, 2015)

La organización de las naciones unidas para la educación, la ciencia y la cultura (UNESCO, 2009) vela por la conservación, el progreso y la difusión del saber, conservando y

protegiendo el patrimonio universal de la sociedad (libros, documentos, obras). Esta organización define en una carta: la preservación del patrimonio digital (Melorose et al., 2015), las directrices para la preservación digital, abarcando los ámbitos: el patrimonio cultural y patrimonio documental. Según la Unesco, el patrimonio documental consta de dos componentes: el contenido informativo y el soporte en el que se consigna. A partir de lo anterior, plantea que dicho patrimonio está conformado por: piezas textuales, no textuales, audiovisuales y documentos virtuales. (Biblioteca Nacional de España, 2015)

Un documento digital nace electrónico o de la digitalización del documento en papel. Las normas ISO proponen soluciones y acciones para garantizar que los documentos digitales se conserven a lo largo del tiempo, afrontando los riesgos de la obsolescencia tecnológica.(Chornet, 2014)

Para garantizar que estos documentos digitales sean exactos, fiables e íntegros con el paso del tiempo, se sugiere la norma ISO 18492: *“La conservación a largo plazo de la información basada en documentos”*(Chornet, 2014), donde se destaca los siguientes criterios:

- Diseñar una estrategia de conservación a largo plazo, abordando la obsolescencia tecnológica e incompatibilidad de HW y SW.
- Transferir la información desde la producción, generadores y receptores, a un sistema de almacenamiento de gestión documental.
- Almacenar en un soporte estable contra la obsolescencia tecnológica
- Asegurar al acceso y la protección de la información, evitando la alteración de la información.
- Automatizar las migraciones
- Conservar la autenticidad de la información en el cambio de formato
- Proteger la información de documentos alterados o perdidos
- Asegurar que los metadatos sean interoperables
- Abordar la dependencia de SW. Migrar a formatos normalizados. La norma ISO18492, recomienda pdf/a-i, .xml, tiff y jpeg.

- Controlar la calidad en el proceso de migración. Análisis del sistema, diseño de interfaz y base de datos.
- Controlar el acceso a la aplicación o SW.
- Controlar el acceso físico.
- Proteger contra pérdida física del depósito de almacenamiento.
- Controlar y supervisar condiciones ambientales.
- Usar metadatos para la conservación de la información, que garanticen la autenticidad, fiabilidad, integridad y disponibilidad. Norma ISO23081 (Metadatos para la gestión de documentos).
- Desarrollar un esquema de metadatos relacionados con la gestión de documentos.
- Valorar por cuánto tiempo se debe conservar los metadatos.
- Almacenar los documentos en formato neutral y los metadatos se recomienda .xml. Norma ISO19005 (Formato de fichero de documentos electrónicos para la conservación a largo plazo).
- Favorecer como formato de conservación el pdf/a.
- Encapsular y empaquetar la información de los documentos de archivos en el momento de la ingesta. Se debe crear un paquete de archivo (AIP) que contiene el documento y sus metadatos, y enviar el paquete del sistema productor al depósito de almacenamiento (SIP). Norma ISO14721 (Sistema de transferencia de información y datos espaciales OAIS).

Entre los esquemas basados en formato .xml para la codificación de metadatos relacionados con la preservación de objetos digitales están METS y PREMIS.

- Convertir y migrar los documentos electrónicos. Existe una planificación, requisitos y procedimientos para preservar la autenticidad, fiabilidad, integridad y disponibilidad de cada documento. Norma ISO13008 (Proceso de migración y conversión de documentos electrónicos).
- Implementar un sistema de gestión de la seguridad de la información para prevenir o evitar la pérdida de documentos digitales. Norma ISO27001 (Sistema de gestión de la seguridad de la información)

Normas ISO

ISO18492: Conservación a largo plazo de la información basada en documentos digitales. (ICONTEC, 2013)

ISO23081: Información y documentación. Procesos de gestión de documentos. Metadatos para la gestión de documentos. (Comité Técnico de Normalización 50 AENOR, 2008)

ISO19005: Formato de fichero de documento electrónico para la conservación a largo plazo. (ICONTEC, 2005)

ISO13008: Información y documentación. Proceso de migración y conversión de documentos electrónicos.

ISO27001: Sistema de Gestión de la Seguridad de la Información – SGSI. (INCONTEC, 2006)

ISO14721: Sistema de transferencia de información y datos especiales. Sistema abierto de información de archivo (OASIS). (Chornet, 2014)

3. Metodología Aplicada

3.1 Metodología del Marco Lógico

El diseño arquitectónico del sistema de información propuesto en este trabajo se fundamentó en los lineamientos y recomendaciones dadas en la normatividad colombiana vigente, en la *“Guía para la Gestión de Documentos y Expedientes Electrónicos”*, (MINTIC, 2017) , en las definiciones y conceptos presentados por los autores Claudia Milena Rodríguez A y Henry Gaitán Gómez en su libro *“Sistemas de Gestión Documental: Guía Práctica para el Diseño e Implementación”*. (Rodríguez A & Gaitán Gómez, 2011), en la Norma ISO 14721:215 *“Modelo de Referencia OAIS: Sistema de Información de Archivo Abierto”*, en los metadatos para la Preservación PREMIS Versión 3, en la Resolución 31 C/34 dada por la UNESCO (UNESCO, 2003), en la norma ISO 23081 la cual contempla la utilización de metadatos bajo los modelos DUBLIN CORE y METS .

Para el manejo de Metadatos se optó por utilizar el modelo PREMIS, debido a que comparado con METS y DUBLIN CORE, este contempla un conjunto más amplio de datos, además, el modelo PREMIS permite ampliar o escalar todas sus entidades con lo cual podemos incorporar dentro de este los datos faltantes utilizados por METS y por DUBLIN CORE.

La metodología utilizada en este proyecto está alineada a metodologías estándar, utilizadas en la industria en cada una de las fases del ciclo de vida del Software. Es así como para las etapas de: Análisis y Diseño se utilizó la metodología propuesta por el Lenguaje Unificado de Modelado “UML” y para la etapa de Desarrollo se utilizó y recomendamos SCRUM – desarrollo ágil.

3.1.1 Análisis

En la primera fase denominada Análisis, se identifican los requerimientos funcionales y no funcionales del sistema de información. Se emplea la metodología UML con definición de los “Casos de Uso”, para describir las acciones que debe realizar el software desde el punto de vista del usuario final.

“La vista de casos de uso modela la funcionalidad de un sujeto (como un sistema) como lo perciben los agentes externos, llamados actores, que interactúan con el sujeto desde un punto de vista particular. Un caso de uso es una unidad de funcionalidad expresada como una transacción entre los actores y el sujeto. El objetivo de la vista de casos de uso es enumerar los actores y casos de uso y mostrar qué actores participan en cada caso de uso. El comportamiento de los casos de uso se expresa mediante vistas dinámicas, particularmente la vista de interacción” (Awaad, Krauss, & Schmatz, 1978)

3.1.2 Diseño

Concluida la fase de análisis y con el insumo de los casos de uso, el proceso se centra en la construcción del diseño arquitectónico del Sistema de Información. Esta etapa es muy importante para generar un diseño arquitectónico que facilite el desarrollo de la solución. La solución debe cumplir con los requerimientos funcionales establecidos por el usuario o la normatividad, ser eficaz y eficiente para garantizar una buena experiencia de usuario final y

que la aplicación optimice el uso de los recursos de la plataforma que la soporta. Para el diseño se utilizó el modelo de clases y el modelo de datos recomendados por UML. (Awaad et al., 1978)

3.1.3 Desarrollo

Es la etapa donde se definen las actividades propias de la programación de software (lenguaje de programación y Framework). Para el desarrollo del proyecto se empleó el Lenguaje de Programación C#, bajo el conjunto de herramientas que presenta Microsoft Visual Studio y la metodología de programación o marco de trabajo llamado: MVC (Modelo – Vista – Controlador) (Camarena Sagredo et al., 2012). En la programación se emplea la metodología de desarrollo ágil: SCRUM.

3.2 Metodología UML

El proceso unificado de desarrollo de software es el conjunto de actividades necesarias para transformar los requisitos de un usuario en un sistema de software; es un marco de trabajo para una gran variedad de software.

El proceso unificado utiliza el lenguaje unificado de modelado UML para preparar todos los esquemas del sistema.

El Lenguaje de modelado unificado (UML) es un lenguaje de modelado visual de propósito general que se utiliza para especificar, visualizar, construir y documentar los artefactos de un sistema de software. Captura decisiones y comprensión sobre los sistemas que deben construirse. Se usa para comprender, diseñar, explorar, configurar, mantener y controlar la información sobre dichos sistemas.(Awaad et al., 1978)

3.3 Metodología para el desarrollo de software

Como metodología de desarrollo, se utilizó: desarrollo Ágil – SCRUM, en donde de manera cíclica se definieron alcances de desarrollo y se verificaron los desarrollos parciales del mismo.

SCRUM es una metodología de desarrollo ágil que tiene como base la idea de creación de ciclos breves para el desarrollo, comúnmente se llaman iteraciones y en SCRUM se llaman Sprints.(Trigas Gallego & Domingo Troncho, 2012)

Las 5 fases del desarrollo ágil comprenden (Ver Ilustración 4): concepto, especulación, exploración, revisión y cierre.



Ilustración 4 – Ciclo de Desarrollo Ágil (Trigas Gallego & Domingo Troncho, 2012)

Concepto: Define de forma general las características del producto y se define el equipo de desarrollo.

Especulación: Se establecen los alcances establecidos para el ciclo a ejecutar. Esta fase se repite en cada iteración y contempla:

- Desarrollar y revisar los requisitos establecidos.
- Mantener la lista de las funcionalidades esperadas.
- Establecer el plan de entrega (Fechas, hitos, entre otras).

Exploración: Se incrementa el producto ya que se le añaden las funcionalidades de la fase de especulación.

Revisión: El equipo revisa todo lo construido y lo compara con el objetivo y alcance propuesta para la iteración

Cierre: Se entregará en la fecha acordada una versión del producto deseado. El cierre no indica que ha finalizado el proyecto, sino que seguirá habiendo cambios, denominados “mantenimiento” y que acercaran cada vez más el software al producto final esperado. (Trigas Gallego & Domingo Troncho, 2012)

En el desarrollo de software es muy difícil congelar los requerimientos establecidos en la fase de análisis por los constantes cambios que se presentan. Las metodologías de desarrollo ágil promueven ciclos cortos de entrega y avances funcionales y periódicos en el desarrollo e implementación del sistema de información. Cada ciclo se conoce como iteración, cada iteración se divide en *scrums* y *sprints*, cada iteración se completa con una entrega con la cual el equipo de desarrollo se retroalimenta con las observaciones dadas por los usuarios, esta metodología de trabajo permite adaptar el desarrollo del software a continuos cambios en la definición de los requisitos, que los usuarios puedan ver de manera rápida y periódica los continuos avances y un trabajo más articulado entre los usuarios y el equipo de desarrollo. (Sinha, 2010).

Los requerimientos de este proyecto se rigen por la normatividad y recomendaciones dadas por el Archivo General de la Nación. La generación de nuevos requerimientos que no estaban previamente establecidos son poco frecuentes, por lo que se puede realizar gran parte del desarrollo del software siguiendo la metodología cascada: Análisis, Diseño, Implementación y Pruebas); sin embargo, el enfoque ágil se utiliza sin ningún problema, y podemos obtener de él, los demás beneficios que éste ofrece, por ejemplo: realizar iteraciones que permitan tener en cuenta los requerimientos no funcionales y el diseño de las vistas o capa de presentación del modelo MVC, con lo cual se garantiza que la experiencia de usuario y la eficacia del software se pueda ver y mejorar desde el inicio del desarrollo.

Para todas las fases a ejecutar en el análisis, diseño y programación del sistema de información documental, se adoptó también, partes del estándar ESA PSS-05-0 I (ESA, 1994).

ESA PSS-05-0 describe los estándares de ingeniería de software que se aplicarán a todos los artefactos software implementados por la Agencia Espacial Europea (ESA). Este estándar requiere que la definición de los requisitos de software y el diseño arquitectónico se

desarrollen en fases separadas. Estas fases terminan con una revisión formal del documento de requisitos de software y el documento del diseño arquitectónico. Las plantillas de documentos proporcionados en ESA PSS-05-0 se basan en los estándares ANSI/IEEE, y están diseñados para satisfacer los requisitos de documentación de todos los proyectos.

3.4 Metodología de Minería de Datos

La minería de datos es una disciplina para explorar la información y obtener nuevos conocimientos a partir de los datos. Se aplica a grandes volúmenes de datos, en todos los sectores y campos, y permite identificar información oculta o implícita, que no es posible obtener mediante métodos estadísticos tradicionales. El proceso de extracción de la información requiere de la aplicación de una metodología estructurada, utilizando técnicas y herramientas existentes.

“La minería de datos (DM, Data Mining) consiste en la extracción no trivial de información que reside de manera implícita en los datos. Es una tecnología con gran potencial para ayudar a las compañías a concentrarse en la información más importante de sus Bases de Información. Las herramientas de Data Mining predicen futuras tendencias y comportamientos, permitiendo en los negocios tomar decisiones proactivas y conducidas por un conocimiento acabado de la información”.(Miñambres, 2007)

En la actualidad se destacan tres metodologías para el proceso de la minería de datos: KDD, CRISP-DM y SEMMA, estos modelos plantean un enfoque sistemático para llevar a cabo el proceso de Data Mining. Comparten la misma estructura por fases, haciendo el proceso iterativo e interactivo.

“Las metodologías SEMMA y CRISP-DM comparten la misma esencia, estructurando el proyecto de Data Mining en fases que se encuentran interrelacionadas entre sí, convirtiendo el proceso de Data Mining en un proceso iterativo e interactivo. La metodología SEMMA se centra más en las características técnicas del desarrollo del proceso, mientras que la metodología CRISP-DM, mantiene una perspectiva más amplia respecto a los objetivos

empresariales del proyecto. Esta diferencia se establece desde la primera fase del proyecto de Data Mining donde la metodología SEMMA comienza realizando un muestreo de datos, mientras que la metodología CRISP-DM comienza realizando un análisis del problema empresarial para su transformación en un problema técnico. Desde ese punto de vista más global se puede considerar que la metodología CRISP-DM está más cercana al concepto real de proyecto, pudiendo ser integrada con una Metodología de Gestión de Proyectos específica que completaría las tareas administrativas y técnicas.” (Montequín et al., 2003)

En un estudio realizado en el año 2007 por la comunidad KDnuggets (Data Mining Community's Top Resource), CRISP-DM se ha convertido en la metodología más utilizada en el desarrollo de proyectos de Data Mining. Este modelo estructura el proceso en seis fases y cada fase es descompuesta en varias tareas generales. Las tareas generales se proyectan a tareas específicas, pero en ningún momento se propone como realizarlas. Es decir, el modelo establece un conjunto de tareas y actividades para cada fase del proyecto pero no especifica cómo llevarlas a cabo (Moine, Haedo, & Gordillo, 2011).

Sus 4 niveles de abstracción (Ver Ilustración 5), están organizados de forma jerárquica y organiza el desarrollo de un proyecto de Data Mining, en una serie de 6 fases (Ver Ilustración 6).

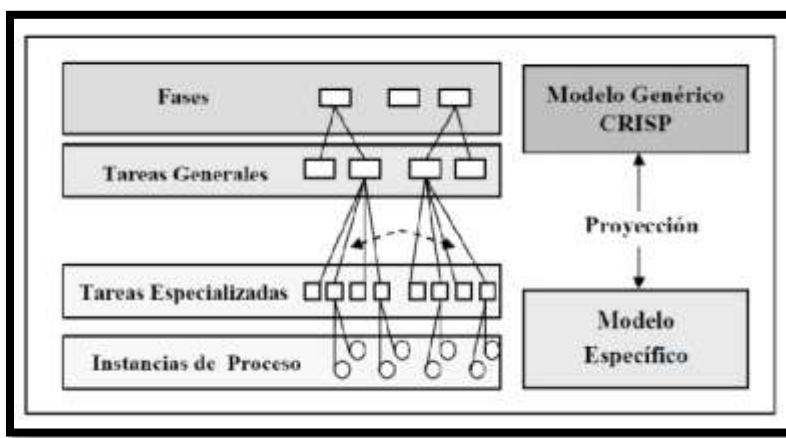


Ilustración 5 – Niveles de abstracción de CRISP-DM (Gallardo, 2009)

Las fases que divide CRISP-DM son:

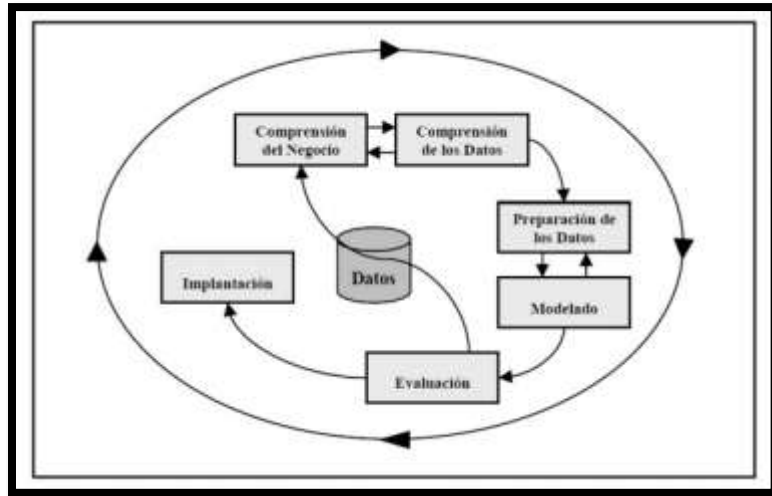


Ilustración 6 – Modelo de proceso CRISP-DM (Gallardo, 2009)

- **Comprensión del problema:** Es entender el problema que se desea resolver, lo que permite la recolección de datos correctos e interpretación de resultados. Las tareas comprenden la determinación del problema, identificar el porqué de la necesidad de usar Data Mining y definir los criterios de éxito.
- **Comprensión de los datos:** Comprende la recolección de los datos, identificación de su calidad y establece las relaciones para definir las primeras hipótesis. Las tareas a desarrollar son: recolección inicial, descripción, exploración y verificación de la calidad de los datos.
- **Preparación de los datos:** Efectuada la recolección de los datos, se procede a prepararlos para adaptarlos a técnicas de Data Mining. Incluye tareas generales como selección de datos, limpieza, integración de diferentes orígenes de datos y cambios de formato. Esta fase está relacionada con la de modelado e interactúan de forma permanente.
- **Modelado:** Se selecciona las técnicas de modelado más apropiadas, se eligen en función de: apropiación al problema, datos adecuados, tiempo adecuado, conocimiento de la técnica y cumplir los requisitos del problema. Después de

modelado se determina un método de evaluación para establecer el grado de bondad de ellos.

- **Evaluación:** Se evalúa el modelo teniendo en cuenta los criterios de éxito del problema. Se puede considerar emplear herramientas para la interpretación de resultados, ejemplo para un modelo de clasificación se usa las matrices de confusión. Si el modelo generado es validado en función de los criterios de éxito se procede con la explotación del modelo. Las tareas de esta fase comprenden: evaluación de resultados, proceso de revisión y determinación de futuras fases.
- **Implementación:** Una vez modelo ha sido construido y validado, se transforma el conocimiento obtenido en acciones dentro del proceso de negocio. Las tareas que se ejecutan en esta fase son: plan de implementación, monitorización y mantenimiento, informe final y revisión del proyecto.

En la fase de pruebas del prototipo de migración se emplea el modelo CRISP-DM expuesto. Este modelo permite profundiza en mayor detalle sobre las tareas y actividades a ejecutar en cada etapa. Partiendo de los resultados obtenidos por las pruebas del prototipo se quiere encontrar patrones y relaciones que aporten nuevo conocimiento para la solución del mismo.

4. Requerimientos Funcionales del Sistema de Información Documental

Se identifican las características básicas de los requerimientos funcionales realizados para el Sistema de Gestión y Administración Documental; para cada uno de estos elementos se anexa su correspondiente documento técnico, el cual explica a detalle cada componente.

Para la documentación de los requerimientos funcionales se utilizó la metodología UML utilizando como herramienta los “Casos de Uso”.

Los casos de uso documentan la funcionalidad completa de un sistema (o una clase) mostrando su interacción con los agentes externos. Esta representación se hace a través de las relaciones entre los actores (agentes externos) y los casos de uso (acciones) dentro del sistema. Los diagramas de casos de uso definen conjuntos de funcionalidades afines que el

sistema debe cumplir para satisfacer todos los requerimientos que tiene a su cargo. Esos conjuntos de funcionalidades son representados por los casos de uso. Se pueden visualizar como las funciones más importantes que la aplicación puede realizar o como las opciones presentes en el menú de la aplicación.

El listado de casos de uso se detalla en el “**ANEXO 1 - CASOS DE USO**”

5. Arquitectura, Modelo de Clases y Modelo de BD del Sistema de Información Documental

5.1 Requerimientos Arquitectónicos y Herramientas de Desarrollo

Se presenta el diseño de la arquitectura del sistema de información propuesto utilizando las herramientas que define UML (Lenguaje Unificado de Modelado), en este caso, se presentará el modelo, diagrama de clases y el diseño físico de la base de datos acompañado del diccionario de datos.

5.1.1 Diseño de Despliegue

El despliegue del sistema de información se realiza sobre los siguientes componentes (Ver Ilustración 7):

- **Servidor de Base de Datos:** Máquina de alta disponibilidad utilizada para hospedar el motor de base de datos relacional (Database Management System - DMBS), el cual puede ser:
 - MS SQL Server.
 - Oracle.
 - PostgreSQL.
 - DB2.

En el DMBS se crea la Base de Datos del Sistema de Información de Gestión Documental.

- **Servidor de Archivos:** Equipo de alta disponibilidad utilizado como repositorio central de los documentos digitales. Todos los documentos digitales se almacenan dentro de directorios estructurados jerárquicamente a partir de un directorio raíz. El sistema de información está diseñado para que los documentos de cada unidad documental se almacenen a partir de un directorio raíz previamente establecido por el usuario CAD, el sistema va creando subdirectorios: por año, por fecha (uno por cada día o fecha), en donde los documentos se van almacenado.
- **Contratos**
 - 2017
 - 01_01_2017
 - Documento1.pdf
 - Documento2.pdf
 - ...
 - 01_02_2017
 - Documento3.pdf
 - ...
- **Servidor de Web:** Servidor en donde se hospedada:
 - La aplicación web de usuario final (Frontend)
 - La aplicación web de usuarios administradores (Backend)
 - La aplicación web que presenta los servicios web (WebAPIs)
- **Servidor LDAP:** Este componente es opcional ya que es solo utilizado cuando se requiere que la aplicación valide a los usuarios del módulo privado contra un servidor LDAP.
- **DM:** Dispositivo móvil, en donde se instala la APP de la aplicación.

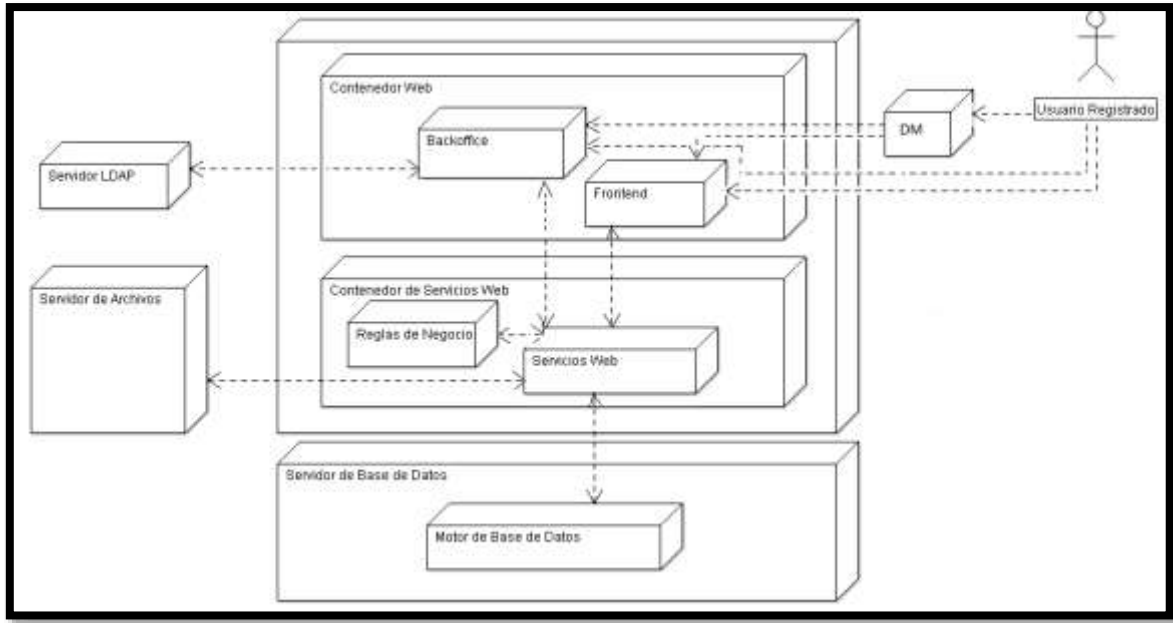


Ilustración 7 – Diagrama de Despliegue

5.1.2 Herramientas de Desarrollo

Para la fase de desarrollo del prototipo funcional del módulo de migración, se utilizaron las siguientes herramientas de programación:

- Microsoft Visual Studio Community, versión 2017. ASP.Net y Lenguaje de Programación C#.
- Se utilizó el patrón MVC (Modelo - Vista - Controlador) proporcionado por Visual Studio(Microsoft, 2009)
- Para la construcción de Dashboards como herramientas de BI y análisis de datos, se utilizó la librería **DevExpress**, versión 18.1.

Se utilizó los requerimientos funcionales y no funcionales expuestos, para el diseño arquitectónico y estructural del sistema de información de gestión y administración documental.

5.2 Diseño y Arquitectura del Sistema de Gestión Documental

El diseño arquitectónico del sistema de información está basado en los requerimientos funcionales y no funcionales establecidos, tiene como objetivo describir los componentes básicos del sistema de software y su combinación interna, permitiendo entender los alcances, funcionalidades y toda la arquitectura interna necesaria para las fases del desarrollo y mantenimiento del sistema. Esta actividad, tiene como producto una serie de documentos técnicos, los cuales presentan diversas perspectivas de la arquitectura del sistema.

Se anexan los documentos técnicos, los cuales poseen todo el diseño arquitectónico del sistema propuesto, a continuación, se hace un resumen de cada uno de los documentos anexos que hacen parte de este documento.

5.2.1 Modelo Lógico

El modelo lógico enseña los elementos estructurales estáticos que componen la solución, se usa UML (Lenguaje Unificado de Modelado). Como herramientas para el modelado lógico se utilizaron: modelo de clases y el modelo de la base de datos (Olga Pons, Nicolás Marín, Juan Miguel Medina, Silvia Acid & Vila, 2005) ; en este último el modelo que aplica es el de Entidad Relación, porque es una base de datos Relacional.

5.2.2 Modelo de Clases

El documento “**ANEXO 2 - MODELO DE CLASES**” presenta toda la documentación de los modelos de clases diseñado para soportar la implementación del sistema de administración documental propuesto. En el diseño se utiliza el patrón MVC (Modelo - Vista - Controlador). Como el alcance del proyecto llega hasta el diseño arquitectónico de la solución propuesta, sólo se modelaron las clases correspondientes a la capa del modelo en dicho patrón; la empresa o persona que desee desarrollar el sistema propuesto deberá completar la arquitectura adicionándole clases, métodos a las ya existentes, con el fin de diseñar la capa

de controladores y programar la capa de vistas que corresponde a las interfaces gráficas que ve el usuario final del sistema de información.

5.2.3 Modelo de Base de Datos

El diseño de la base de datos parte de la creación de los siguientes esquemas:

- **ARCHIVO:** Almacena las tablas que tienen relación con la clasificación documental, tablas de retención y valoración documental, índices de búsqueda de documentos y expedientes, y el almacenamiento de los documentos y expedientes.
- **TRAMITES:** Contiene las tablas que permiten la configuración de los procesos por los que fluyen los documentos dentro de la organización (Workflow); igualmente las tablas que permiten almacenar toda la información relacionada con los trámites que se generan a partir de los procesos configurados.
- **SEGURIDAD:** Contiene las tablas destinadas a la administración y auditoría de la seguridad del sistema de información y la de los documentos y expedientes electrónicos creados por los usuarios.

En el documento “**ANEXO 3 - MODELO FÍSICO DE DATOS**” se presenta la relación de las tablas que conforman el diseño de la base de datos de la aplicación, con el nivel de detalle de todos los objetos que componen el modelo; contempla el diagrama entidad relación y diccionario de datos.

El primer grupo de tablas son las que permiten la clasificación documental, está conformada por: **SERIE_DOCUMENTAL**, **SUBSERIE_DOCUMENTAL**, **UNIDADES DOCUMENTALES**, **DISPOSICION_FINAL**.

La tabla **SERIE_DOCUMENTAL** es utilizada para almacenar las series documentales creadas por el CAD y aprobadas por el comité directivo de cada entidad; en ella se almacena la clave primaria, el nombre de la serie documental, la descripción (opcional) y un campo utilizado en la mayoría de las tablas que conforman el modelo llamado **B_ACTIVO**, el cual es utilizado para habilitar o deshabilitar en cualquier momento (se asemeja a un borrado

lógico). Posee una relación de uno a muchos con la tabla SUBSERIE_DOCUMENTAL, porque una serie documental puede agrupar varias subseries documentales.

Cada SUBSERIE DOCUMENTAL agrupa a su vez una o varias UNIDADES DOCUMENTALES. Una Unidad Documental puede ser clasificada como simple y se define como tipo documental, cuando está constituida por un solo documento, o compleja, cuando está constituida por varios documentos relacionados y que conforman un expediente. (Archivo General de la Nación, 2001).

La Tabla UNIDAD_DOCUMENTAL está diseñada con el fin de almacenar las unidades documentales pertenecientes a cada Sub Serie Documental.

A este conjunto de tablas, se le adiciona la tabla DISPOSICION_FINAL, la cual es utilizada para almacenar el listado de los posibles destinos a los que están sometidos los documentos, una vez cumplen su tiempo en el *archivo histórico*.

Los tipos de disposición final son los que definió la Armada Nacional de Colombia en el año 2014. (Reynaldo Antonio Espinosa, 2014)

- Conservación Total: Se conservan totalmente, aquellos documentos que adquieren con el tiempo, valores secundarios, es decir se convierten en fuentes de consulta para la investigación, la ciencia y la cultura.
- Microfilm: Esta técnica que permite registrar fotográficamente documentos como pequeñas imágenes en películas de alta resolución, se aplica a aquellos documentos que se van tanto a conservar totalmente, como a seleccionar con el fin de garantizar la consulta de la información en otro soporte, para asegurar la preservación del documento original.
- Selección: Este procedimiento se aplica para aquellas series y/o subseries documentales que no ameritan conservarse totalmente, pero poseen valor informativo que se repite o se encuentra registrado en otras series y/o subseries documentales. Una vez se efectuó la selección, implícitamente se está determinando eliminación sobre la documentación no seleccionada.

- Eliminación: Se aplica para aquellos documentos que pierden sus valores primarios y no adquieren valores secundarios. La eliminación es un proceso irreversible, por lo que esta decisión debe estar muy bien sustentada.

Otras tablas relacionadas con la unidad documental son: retención y el tipo documental, que permiten almacenar periódicamente la información de las tablas de retención documental. Ver **“ANEXO 3 - MODELO FÍSICO DE DATOS”**.

Cada entidad, define o actualiza periódicamente sus Tablas de Retención Documental (TRD), estas son un listado de series y sus correspondientes tipos documentales a las cuales se asigna el tiempo de permanencia en cada etapa del ciclo vital de los documentos. (Archivo General de la Nación, 2001)

La Tabla TABLAS_RETENCION está diseñada para almacenar el historial de las versiones de las tablas de retención documental establecidas por la entidad a través de su CAD. El modelo cuenta con una tabla llamada UNIDAD_TABLARET que se encuentra relacionada con las tablas UNIDAD_DOCUMENTAL y TABLAS_RETENCION y que es utilizada para establecer las unidades documentales que pertenecen a cada una de las versiones de las tablas de retención documental.

La tabla TIPO_DOCUMENTAL permite configurar los tipos de documentos que están asociados a una Unidad Documental Simple.

Como se deben asignar índices de búsqueda a los documentos con el fin de que por medio de ellos los usuarios puedan realizar las consultas de los mismos documentos, se diseñó un conjunto de tablas relacionadas con la tabla UNIDAD_DOCUMENTAL, las cuales cumplen dicha función.

Cada Unidad Documental posee sus propios índices de búsqueda, es decir, los índices de búsqueda de un documento son los que están asociados a la unidad documental a la que ellos pertenecen.

La tabla LISTADO_INDICES almacena una serie de índices, cada uno de los cuales puede ser utilizado en diferentes unidades documentales y/o expedientes electrónicos. La tabla

INDICES_UNIDADDOC está relacionada con las tablas UNIDAD_DOCUMENTAL y LISTADO_INDICES y se diseñó para establecer los Índices de búsqueda que están asociados a cada Unidad Documental.

Se contemplaron dos posibles fuentes u orígenes de datos asociados a cada uno de los listados, una fuente representa valores almacenados en la tabla DETALLE_LISTADO y son administrados desde la misma aplicación de Gestión y Administración Documental. Existe una segunda fuente de valores para un listado y son datos obtenidos desde una consulta o Query a una base de datos externa y de la que se tiene configurada y almacenada su cadena de conexión y la consulta misma desde la cual se obtienen dichos valores.

La tabla LISTADOS permite la configuración de cada uno de los listados que se desean utilizar como tipos de datos de los índices.

La tabla DETALLE_LISTADO se pensó y modeló con el fin de que los valores de un listado sean administrados por la misma aplicación y se almacenen en la misma base de datos del sistema, específicamente en esta tabla.

Existen documentos que nacen digitalmente y que los usuarios del sistema de información crean, radican, tramitan, firman con certificado digital y almacenan en el sistema, los cuales son creados a partir de plantillas de documentos preconfiguradas por usuarios con el rol o perfil correspondiente.

Para poder configurar plantillas de documentos, se modelaron las siguientes tablas. Ver **“ANEXO 3 - MODELO FÍSICO DE DATOS”**

La tabla PLANTILLAS, almacena la información básica de cada una de las plantillas, a partir de las cuales se crearán documentos digitales cuyo origen no proviene de un soporte físico. Como todos los documentos deben reposar en el CAD y estos a su vez deben ser clasificados de acuerdo con las Tablas de Retención Documental e indexados para que puedan ser luego consultados, se hace necesario asignarle a cada plantilla el Id de la Unidad Documental a la que pertenecerán todos los documentos que sean creados a partir de ella, con esto ya se sabe cuáles son los índices que deben tener asociados los documentos ya la unidad documental posee un conjunto de índices relacionados. Adicionalmente, en el diseño se contempló la

posibilidad, de que los usuarios le creen otros campos o datos asociados a la plantilla independientes de los índices de búsqueda de la Unidad documental asignada, dado que, en ocasiones, la plantilla puede tener muchos más datos que se pueden auto complementar, que no están contemplados en los índices de búsqueda.

Las tablas hasta aquí explicadas se utilizan para la configuración inicial del sistema o para el mantenimiento de dicha configuración, ahora veremos las tablas que se utilizan para almacenar, radicar e indexar los documentos que ingresan al CAD día a día.

Con el fin de que el sistema de Gestión y Administración documental permitiera el manejo y uso de Expedientes Electrónicos, se modelaron el grupo de siguientes tablas.

Para el manejo de los Préstamos de los expedientes físicos se integran las tablas: DOCUMENTOS, DETALLE_PRESTAMOS, EXP_CARPETA, EXP_ANEXOS, DETALLE_PRESTAMOS, PRESTAMOA, TIPO_PRESTAMO.

6. Prototipo Funcional del Módulo de Migración del SGD

Partiendo de la arquitectura propuesta para la programación e implementación de un Sistema de Gestión y Administración Documental, se presenta la arquitectura, el funcionamiento y los programas fuentes de un módulo que permite realizar tareas de migración de documentos y que está basado en el estándar PREMIS.

Todos los programas fuentes de este prototipo de migración se anexan a este documento y pueden ser utilizados y modificados según las necesidades de cada implementación, adicionalmente se anexa un documento llamado:

“ANEXO 4 – MODELO DE CLASES MÓDULO DE MIGRACIÓN PREMIS” en el cual se explica al detalle las clases que lo componen.

6.1 Arquitectura del Módulo de Migración

El módulo de migración se diseñó y programó utilizando el patrón de diseño **Modelo–Vista–Controlador**.

“Un patrón recomendable para el desarrollo del software para aplicaciones Web es el Modelo Vista Controlador (MVC), éste considera separar en tres elementos o capas todo el proyecto, siendo: la lógica de control (saber qué elementos tiene el proyecto y qué hacer, pero no cómo se implementó), la lógica de negocio (saber cómo se desarrolla la aplicación) y la lógica de presentación (saber cómo interactúa el usuario con la aplicación)” (Camarena Sagredo et al., 2012).

Se utilizó el lenguaje de Programación C# de Microsoft en el IDE de Desarrollo Microsoft Visual Studio 2017 en su versión Community.

En la Ilustración 8 se presenta la estructura de la solución creada en Visual Studio.

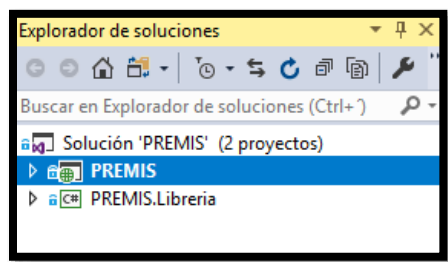


Ilustración 8 – Proyectos de la solución

La solución posee dos proyectos:

- **PREMIS:**

Este proyecto es del tipo *Aplicación web ASP.NET (.NET Framework)* utilizando el patrón MVC.(Camarena Sagredo et al., 2012). Se desarrolló el FrontEnd o aplicación web que el usuario utiliza para realizar sus procesos de Migración. La estructura del proyecto (carpetas y subcarpetas) se detalla en la Ilustración 9.

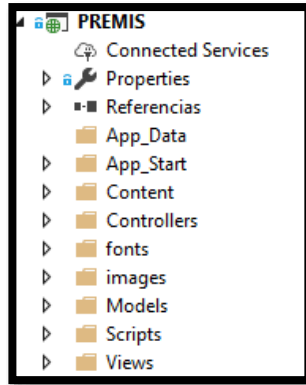


Ilustración 9 – Estructura del Proyecto PREMIS

Se pueden observar las siguientes carpetas en la estructura del proyecto:

- **Models:** En la cual se ubican los modelos según el patrón de diseño MVC.
- **Controllers:** En ella encontramos los Controladores, de acuerdo con MVC.
- **Views:** Utilizada para almacenar las Vistas, tal como lo establece el patrón de diseño MVC.

- **PREMIS.Libreria:**

Este proyecto es una Biblioteca de Clases (.Net Framework), y posee el diseño de la clase PREMIS la cual mapea todos los métodos y sus propiedades de acuerdo con el estándar PREMIS versión 3.

Con el fin de poder realizar un proceso de migración documental y con ello validar el modelo con la generación de indicadores y estadísticas sobre los documentos migrados; se creó una base de datos llamada “Documental” (Ver Ilustración 10). Esta base de datos está basada en el modelo de base de datos propuesto para el Sistema de Gestión y Administración Documental (ya documentada), pero sólo contiene las tablas mínimas necesarias para emular el proceso de migración y no se tiene en cuenta las tablas para el flujo electrónico de documentos, ni las tablas relacionadas con expedientes electrónicos y préstamos, ni tampoco las tablas del esquema de seguridad.



Ilustración 11 – Modelo Entity Framework

6.2 Diseño de los Formularios de la Aplicación

En éste numeral se explican los diseños de los formularios que componen el módulo de migración.

6.2.1 Layout.cshtml

Este componente corresponde al layout o plantilla principal desde la cual todos los formularios de la aplicación heredan su diseño.

Este archivo llamado Layout.cshtml (Ver Ilustración 12), es un archivo en formato html y define la cabecera de las vistas, la cual presenta el menú de opciones de la aplicación. Posee las etiquetas html que definen el pie de página de las vistas

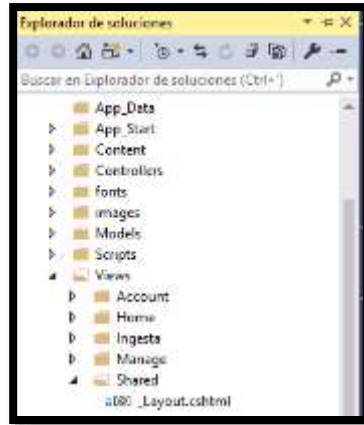


Ilustración 12 – Plantilla principal (layout)

6.2.2 Índice

Índice es la vista que reproduce el formulario inicial y principal de la aplicación tal como se muestra en la Ilustración 13:



Ilustración 13 – Formulario Inicial del Prototipo de Migración

El controlador que por defecto ejecuta la aplicación se llama *HomeController.cs*, esto se encuentra establecido en el archivo *RouteConfig.cs* (Ver Ilustración 14).

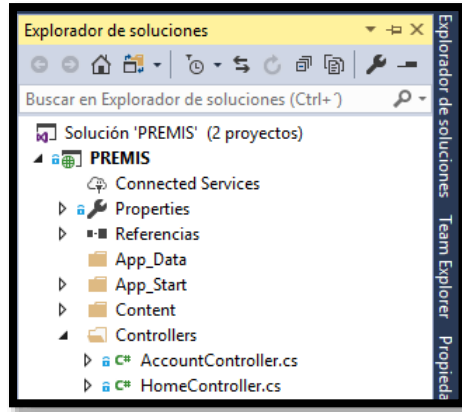


Ilustración 14 – Ubicación archivo HomeController.cs

Este controlador, referencia por medio de sus métodos cuatro vistas:

- **Índex:** Vista que presenta el contenido del formulario principal del módulo de migración.
- **About:** Vista que presenta el formulario *Acerca de*
- **Contact:** Vista que presenta el formulario de *Contactos*.

La vista Índex (Ver Ilustración 15) en su diseño muestra en la parte superior un objeto de tipo carrusel el cual va mostrando un conjunto de banners cada uno de los cuales posee un botón link a una página guía.



Ilustración 15 – Banners del Carrusel

Por último tal como se muestra en la Ilustración 16, la vista presenta dos opciones en su parte inferior:



Ilustración 16 – Opciones adicionales de la vista Índice

La *primera* opción, permite invocar la vista ayuda asociada al controlador *ProcesoController*, es utilizada para mostrar una página de ayuda al usuario que desea realizar un proceso de migración.

La segunda opción está relacionada con la vista llamada *DocumentosSinMetadatos* perteneciente al controlador *IngestaController* y es la que muestra al usuario la opción y funcionalidad disponible para que se pueda realizar el proceso de migración documental.

6.2.3 Ingesta – Migración de Documentos

El formulario detallado en la Ilustración 17, permite realizar el proceso de migración de un conjunto de documentos que no poseen metadatos y que se encuentran almacenados en un directorio, a una base de datos similar a la propuesta para el sistema de gestión y administración documental, incorporándole además a los documentos, el archivo de metadatos en formato PREMIS. El algoritmo desarrollado también realiza una copia de los documentos ya migrados a un directorio de destino y copia adicionalmente de manera redundante el archivo PREMIS de cada uno de los documentos, redundante porque ya cada documento incorpora en su PDF el XML PREMIS.

Preservación Documental Inicio Acerca de Contacto Registrarse Iniciar sesión

Documentos Sin Metadatos

Este procedimiento se encarga de acondicionar un conjunto de documentos almacenados en una estructura jerárquica de directorios, los procesa y los almacena en otra estructura de directorios.

Los documentos de entrada pueden estar creados en diferentes formatos (*.Pdf, *.Tif, *.Doc) y el proceso los entrega en formato (*.Pdf/A); adicionalmente los documentos a la salida del proceso, incorporan un archivo XML, que almacena la Metadatos de acuerdo con el estándar PREMIS versión 3

Archivo Plano PREMIS

Examinar...

Directorio de Origen

Directorio de Destino

Iniciar Proceso...

Documentos Migrados :

© 2018 - Módulo de Migración PREMIS

Ilustración 17 –Formulario para realizar el proceso de Migración

El controlador asociado a la vista: *IngestaController.cs* posee los métodos especializados en realizar la migración de los documentos.

Los parámetros de este método son:

- Directorio Origen: Especifica la ruta del directorio inicial en donde se encuentran los documentos que se desean migrar.
- Directorio Destino: Especifica la ruta del directorio final en donde se almacenan los documentos una vez migrados.
- Archivo Plano PREMIS: Archivo de Texto con el formato establecido y que posee la información de los metadatos básicos de PREMIS que será anexada a cada documento migrado.
- Nro. Documentos Migrados: Parámetro por referencia que se utiliza para que el procedimiento devuelva el número total de documentos que fueron migrados.

Instrucciones del método MigrarDocumentosSinPREMIS

- a) Lleva a una variable de tipo Stream el arreglo de bytes que posee el documento que ingresa por el parámetro ArchivoPlanoPREMIS.

- b) Recorre en un ciclo while cada una de las líneas o filas del documento almacenado en la variable stream.
- c) Por cada fila leída en el ítem anterior llama a la función ProcesarRegistro, función encargada de tratar y procesar cada uno de los documentos.

6.2.3.1 Clase PDF

El proceso de ingesta utiliza la clase llamada PDF, ubicada en la librería de Clases PREMIS.Libreria, esta clase posee los métodos necesarios para obtener la información necesaria para los metadatos PREMIS y que se puede extraer de forma automática de cada uno de los documentos que son migrados, adicionalmente posee un método que permite Insertar el archivo XML PREMIS dentro de un documento PDF.

Los métodos que posee la clase PDF son:

- InsertXMLPremisToPDF: Método que permite Insertar el archivo XML PREMIS dentro de un documento PDF.
- GetNumberPages: Obtiene el número de páginas que posee el documento especificado en el parámetro de entrada
- GetVersion: Retorna la Versión de la aplicación que creó el documento
- GetDateCreated: Este método retorna un dato de tipo *DateTime* el cual especifica la fecha de creación del documento especificado en el parámetro de entrada.
- GetAplication: Este método retorna un dato de tipo string el cual especifica el nombre de la aplicación que creó el documento especificado en el parámetro de entrada.

6.3 Descripción Funcional

Hace referencia a la operación del módulo de migración. Ver “**ANEXO 6 – MANUAL DE USUARIO DEL PROTOTIPO DE MIGRACIÓN**”.

7. Pruebas del Prototipo Funcional de Migración

Para poner a prueba el módulo de migración documental, se realizó un proceso de migración de 200 documentos, con esta muestra podemos evidenciar el comportamiento del algoritmo, su rendimiento y la eficacia de los resultados esperados.

7.1 Plan de Pruebas

Para ejecutar la prueba del módulo se debieron llevar a cabo los siguientes pasos previos:

7.1.1 Preparar la Base de Datos

Los documentos originales ubicados en el directorio de origen deben ser procesados incorporándoles el archivo de metadatos PREMIS, adicionalmente, una vez procesados deben ser llevados también a la base de datos del sistema de información propuesto como solución a una herramienta de administración y gestión documental.

Para realizar el proceso de ingesta en la base de datos, esta debe ser previamente configurada y puesta a punto en sus tablas maestras y de configuración.

7.1.1.1 Series Documentales

En la Ilustración 18 se detallan las Series Documentales creadas y que corresponden a: oficios, actas y contratos.



The image shows a screenshot of a database application window titled 'DESKTOP-490MBQ...ERIE_DOCUMENTAL'. It displays a table with four columns: 'Id_Serie', 'S_Nombre', 'S_Descripcion', and 'B_Activo'. The table contains three rows of data: 'OFICIOS', 'ACTAS', and 'CONTRATOS', all with 'True' in the 'B_Activo' column. A fourth row shows 'NULL' values for all columns.

	Id_Serie	S_Nombre	S_Descripcion	B_Activo
1		OFICIOS	OFICIOS	True
2		ACTAS	ACTAS	True
3		CONTRATOS	CONTRATOS	True
4	NULL	NULL	NULL	NULL

Ilustración 18 – Series Documentales

7.1.1.2 Sub Series Documentales

En la Ilustración 19 se detallan Sub Series Documentales creadas y que corresponden a: oficios despachados, oficios recibidos, actas de juntas, actas de comité, contratos de arrendamiento, contratos de concesión y contratos generales.

Id_SubSerie	Id_Serie	S_Nombre	S_Descripcion	B_Activo
1	1	OFICIOS DESPACHADOS	NULL	True
2	1	OFICIOS RECIBIDOS	NULL	True
3	2	ACTAS DE JUNTA	NULL	True
4	2	ACTAS DE COMITE	NULL	True
5	2	ACTAS DE REUNIONES	NULL	True
6	3	CONTRATOS DE ARRENDAMIENT	NULL	True
7	3	CONTRATOS DE CONCESION	NULL	True
8	3	CONTRATOS GENERALES	NULL	True
NULL	NULL	NULL	NULL	NULL

Ilustración 19 – Subseries Documentales

7.1.1.3 Dependencias

En la Ilustración 20 se detalla la dependencia creada y que corresponde al CAD (Centro de administración documental).

Id_Dependencia	S_Nombre	S_Descripcion	S_Codigointer	B_Estado
1	CAD	CENTRO DE ADMINISTRACIÓN DOCUMENTAL	1	True
NULL	NULL	NULL	NULL	NULL

Ilustración 20 – Dependencias

7.1.1.4 Unidades Documentales

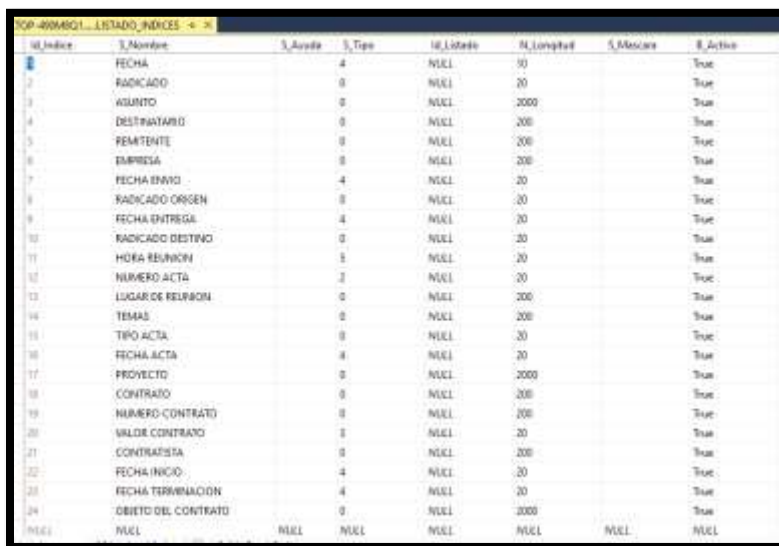
En la Ilustración 21 se detallan las Unidades Documentales creadas:

Id_UnidadDe...	Id_SubSerie	S_Nombre	S_Ruta	B_Definitivo	N_TiempoGest	N_TiempoCen...	N_Ti...
1	1	COMUNICACIONES OFICIALES DESPACHADAS	d:\Documental\Documentos	False	3	30	120
2	2	COMUNICACIONES OFICIALES RECIBIDAS	d:\Documental\Documentos	False	5	30	120
3	3	CORRESPONDENCIA	d:\Documental\Documentos	False	3	30	120
4	3	ACTAS DE JUNTA	d:\Documental\Documentos	False	5	30	120
5	4	ACTAS DE COMITE DE DIRECCION	d:\Documental\Documentos	False	5	30	120
6	4	ACTAS DE COMITE DE ARCHIVO	d:\Documental\Documentos	False	5	30	120
7	4	DE COMITE DE CONTRATACION Y COMPRAS	d:\Documental\Documentos	False	5	30	120
8	5	ACTAS DE REUNIONES	d:\Documental\Documentos	False	5	30	120
9	6	CONTRATOS DE ARRENDAMIENTO	d:\Documental\Documentos	False	5	30	120
10	7	CONTRATOS DE CONCESION	d:\Documental\Documentos	False	5	30	120
11	8	CONTRATOS DE OBRA PUBLICA	d:\Documental\Documentos	False	5	30	120
12	8	CONTRATOS DE PRESTACION DE SERVICIOS PROPIOS	d:\Documental\Documentos	False	5	30	120
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Ilustración 21 – Unidades Documentales

7.1.1.5 Listados de índices

En la Ilustración 22 se detalla los 24 índices creados, junto con sus propiedades:



Id_Indice	S_Nombre	S_Auxiliar	S_Tipo	Id_Listado	N_Longitud	S_Mascara	S_Activo
1	FECHA		4	NULL	30		True
2	RADIADO		0	NULL	20		True
3	ASUNTO		0	NULL	2000		True
4	DESTINATARIO		0	NULL	200		True
5	REMITENTE		0	NULL	200		True
6	EMPRESA		0	NULL	200		True
7	FECHA ENVIO		4	NULL	20		True
8	RADIADO ORIGEN		0	NULL	20		True
9	FECHA ENTREGA		4	NULL	20		True
10	RADIADO DESTINO		0	NULL	20		True
11	HORA REUNION		0	NULL	20		True
12	NUMERO ACTA		2	NULL	20		True
13	LUGAR DE REUNION		0	NULL	200		True
14	TEMAS		0	NULL	200		True
15	TIPO ACTA		0	NULL	20		True
16	FECHA ACTA		4	NULL	20		True
17	PROYECTO		0	NULL	2000		True
18	CONTRATO		0	NULL	200		True
19	NUMERO CONTRATO		0	NULL	200		True
20	VALOR CONTRATO		0	NULL	20		True
21	CONTRAPOSTA		0	NULL	200		True
22	FECHA INICIO		4	NULL	20		True
23	FECHA TERMINACION		4	NULL	20		True
24	OBJETO DEL CONTRATO		0	NULL	2000		True
25	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Ilustración 22 – Listado de Índices

Para el campo *S_Tipo* se establecieron los siguientes valores por defecto

- 0: Alfanumérico
- 1: Entero
- 2: Entero Largo
- 3: Decimal
- 4: Fecha
- 5: Hora
- 6: Listado
- 7: Booleano

7.1.1.6 Índices por Unidad Documental

En la Ilustración 23 le fueron asignados los índices correspondientes a cada Unidad Documental:

Id_Indice/UnidadDoc	Id_UnidadDocumen...	Id_Indice	B_Obligatorio	N_Order
1	1	1	True	1
2	1	2	True	2
3	1	3	True	3
4	1	4	True	4
5	1	5	True	5
6	1	6	True	6
7	2	7	True	1
8	2	8	True	2
9	2	9	True	3
10	2	10	True	4
11	2	3	True	5
12	2	5	True	6
13	2	4	True	7
14	2	6	True	8
15	3	3	True	1
16	3	1	True	2
17	3	4	True	3
18	3	3	True	4
19	4	1	True	1
20	4	10	True	2
21	4	12	True	3
22	4	13	True	4
23	4	14	True	5
24	5	1	True	1

Ilustración 23 – Listado de Índices con su Unidad Documental

7.1.1.7 Tablas de Retención Documental

Se creó una Tabla de Retención Documental y se definió como la TRD vigente, tal como se observa en la Ilustración 24:

Id_TablaReten...	Id_Dependencia	S_Nombre	S_Descripcion	D_InicioVigencia	D_Actualizacion	B_Vigente
2	1	TRD	TABLA DE RETENCIÓN DOCUMENTAL	2018-01-20 00:00:00.0	2018-01-20 00:00:00.000	True
NULL	NULL	NULL	TRD	NULL	NULL	NULL

Ilustración 24 – Tabla de Retención Documental

7.1.1.8 Unidades Documentales por Tablas de Retención

Fueron asignadas las Unidades Documentales a la Tabla de Retención Documental vigente (Ver Ilustración 25).

Id_Unidad_TablaRet	Id_TablaRetencion	Id_UnidadDocumental
1	2	1
2	2	2
3	2	3
4	2	4
5	2	5
6	2	6
7	2	7
8	2	8
9	2	9
10	2	10
11	2	11
12	2	12
NULL	NULL	NULL

Ilustración 25 – Unidades de la TRD Vigente

7.1.2 Preparar los Documentos

Los documentos que se desean migrar deben pasar un proceso de verificación y homologación según lo propuesto por el estándar OAIS, en este proceso se verifican entre otros aspectos que el archivo:

- No esté corrupto o dañado.
- Posea un formato válido.
- Esté libre de virus o amenazas.
- Sea íntegro y seguro.

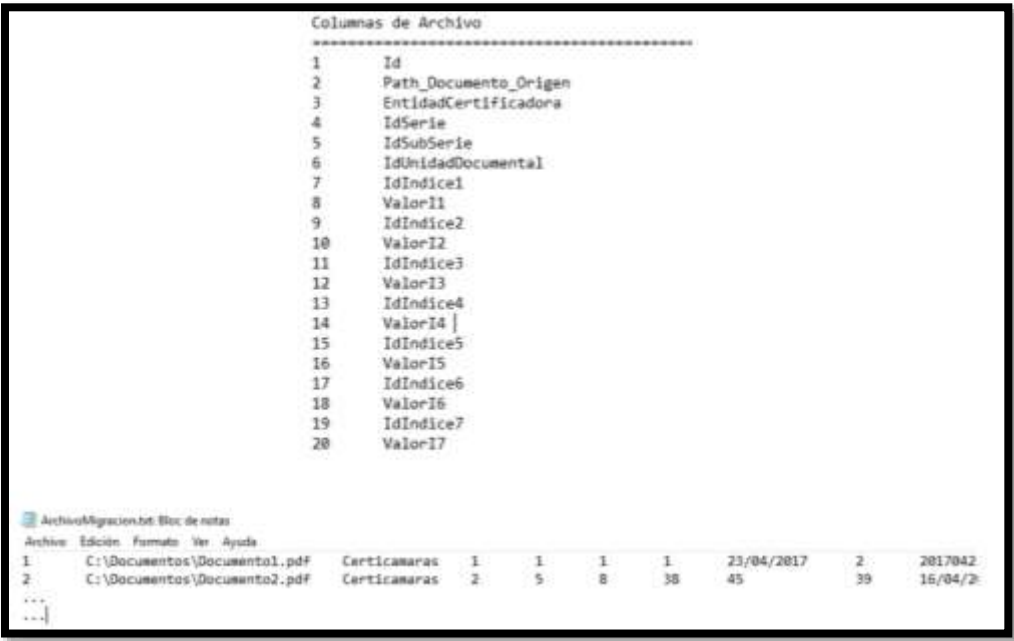
Y se copian todos en un directorio, establecido como el directorio de origen.

7.1.3 Elaborar archivo plano con la información de los documentos a migrar

En el proceso de migración, el programa puede establecer mucha información propia del archivo y necesaria para el archivo de metadatos PREMIS de manera automática, por ejemplo: número de páginas, tamaño del archivo, tipo de archivo, información relacionada con certificados, fecha de creación, última fecha de actualización, extensión, entre otros; pero mucha de la información requerida por PREMIS no puede ser deducida por el algoritmo o

programa de migración, debido a esto es necesario que al momento de realizar el proceso de ingesta, al programa se le deba entregar información adicional relacionada con cada uno de los documentos que se desean migrar, esta información puede ser suministrada por diferentes medios: en base de datos, en archivos XML, en archivo planos, entre otros.

Para nuestro prototipo funcional del módulo de migración se optó por entregarle al programa un archivo de texto plano con las siguientes columnas (Ver Ilustración 26), separadas por el carácter de tabulación:



Columns de Archivo

1	Id
2	Path_Documento_Origen
3	EntidadCertificadora
4	IdSerie
5	IdSubSerie
6	IdUnidadDocumental
7	IdIndice1
8	Valor11
9	IdIndice2
10	Valor12
11	IdIndice3
12	Valor13
13	IdIndice4
14	Valor14
15	IdIndice5
16	Valor15
17	IdIndice6
18	Valor16
19	IdIndice7
20	Valor17

1	C:\Documentos\Documento1.pdf	Certicamaras	1	1	1	1	23/04/2017	2	2017042
2	C:\Documentos\Documento2.pdf	Certicamaras	2	5	8	38	45	39	16/04/2017

Ilustración 26 – Información de la TRD y valor de los Índices de cada documento

Este archivo incluye la información correspondiente a las TRD, y el valor de los índices de cada documento de acuerdo con los índices establecidos para cada unidad documental.

7.1.4 Preparar directorio destino

Se debe crear un directorio en el servidor de archivos como ruta o path raíz a partir de la cual el programa de migración irá colocando los documentos una vez procesados.

7.1.5 Configurar cadena de Conexión a la Base de Datos

Se debe establecer en el archivo web.config de la solución la cadena de conexión a la base de datos

7.1.6 Ejecutar Módulo de Migración

Una vez iniciado el proceso, el programa invoca para cada fila del archivo plano especificado en los parámetros de entrada la función y realiza las siguientes acciones:

- Instancia un nuevo objeto de la clase *LogMigracion* la cual mapea a la tabla correspondiente en la base de datos y que fue diseñada para llevar un log de auditoría del proceso de migración.
- A la propiedad ***HoraInicio*** del objeto *log* le asigna la fecha y hora actual del sistema.
- Crea un arreglo de tipo *string* utilizado para almacenar cada uno de los valores que poseen las columnas de la fila que se está tratando. (Todos los datos que posee el archivo texto para el documento correspondiente a la fila dada).
- Le asigna las propiedades Archivo y Fila al objeto *log*
- Instancia el objeto modelo de la librería *PREMIS*
- Instancia objeto llamado objetoPremis de la clase PREMIS
- Al objeto *objetoPremis* se le asignan valores correspondientes a los metadatos que se desean almacenar.
- Inserta en el objeto los valores extendidos al Modelo PREMIS
- Almacena en los metadatos información relacionada con el repositorio y con el certificado digital del documento
- Almacena información relacionada con el ambiente de Hardware y Software en donde se hospeda y se procesa el documento
- Almacena información relacionada con el evento de Migración
- Crea el archivo PREMIS
- Adiciona el documento PREMIS al documento de salida
- Ingresar los datos faltantes al objeto log y guarda el registro en la base de datos.

7.1.7 Analizar resultados obtenidos en el log de migración.

Para realizar el proceso de Migración, se creó adicionalmente dos tablas en la base de datos LOGMIGRACION Y MIGRACION_LAYOUT. Ver “ANEXO 4 – MODELO DE CLASES MODULO DE MIGRACION PREMIS”

La primera tabla llamada *LogMigracion* se utiliza para registrar la información relacionada con cada uno de los documentos migrados, para cada documento se almacena: el número de la fila en el archivo plano de migración, el nombre del archivo origen, hora con segundos en la que se inició el proceso con el documento, la hora de finalización con segundos del proceso de migración con el documento, un dato llamado estado de tipo booleano que establece si el documento fue migrado satisfactoriamente (true) o (false) de lo contrario y un campo llamado *ErrorM* utilizado para almacenar el mensaje de error producido por la aplicación en el evento de que el documento no pudo ser migrado. Una segunda tabla llamada *Migracion_Layout* almacena por cada documento migrado los dos archivos en formato arreglo de bytes (El documento PDF/A que incluye el XML PREMIS y el archivo XML PREMIS por separado).

Igualmente, el programa de migración sube a la base de datos la información de los documentos migrados en las estructuras de datos definidas para ello.

7.2 Análisis e interpretación de resultados

7.2.1 Resultados almacenados en las tablas de Auditoría

Al realizar el proceso de Migración, con un directorio que contiene 200 documentos, las tablas quedaron tal como se observa en la Ilustración 27:

Id	Fila	Archivo	FechaInicio	FechaFin	Estado	ErrorM
121	119	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
124	120	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
125	121	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
126	122	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
127	123	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
128	124	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
129	125	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
130	126	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
131	127	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
132	128	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
133	129	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
134	130	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
135	131	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
136	132	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
137	133	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
138	134	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	False	El Archivo : E:\Maestra\SistemaDocu...
139	135	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
140	136	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
141	137	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
142	138	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
143	139	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
144	140	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
145	141	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
146	142	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL
147	143	E:\Maestra\Sistema...	2018-06-11 13:0...	2018-06-11 13:0...	True	NULL

Ilustración 27 – Resultados de la Tabla LogMigracion

En ella se observan los 200 registros, uno de los cuales no pudo ser migrado (fila 134), debido a que la ruta estaba mal definida en el archivo plano de migración.

La Tabla *Migracion_Layout* almacenó 199 registros (Ver Ilustración 28), con un documento no se migró.

Id	S_LayoutDocu...	S_LayoutMetadato...
16_Levont		
34b4c68-f6b0-41d2-8518-011205d295d3	<Binary data>	<Binary data>
297a1b6-ca0b-40cd-a048-0177ccaf90	<Binary data>	<Binary data>
7758433e-ecf1-4a25-8c32-020e673488a	<Binary data>	<Binary data>
b6bdf28b-ecf2-4f42-afef-029b6b834b7	<Binary data>	<Binary data>
b13b72dc-bf43-4989-8478-028bd873b774	<Binary data>	<Binary data>
c0d5483c-397a-4345-bd37-094835d728e1	<Binary data>	<Binary data>
f5bec51e-a768-4e6d-8d1a-0358cc072b46	<Binary data>	<Binary data>
ad80a15-220a-4b2d-a25a-94350c306912	<Binary data>	<Binary data>
d67af50c-3a68-483f-89f0-04e1a9b9e2	<Binary data>	<Binary data>
f5922afe-f52c-402c-bf3a-06299be9d8ba	<Binary data>	<Binary data>
25f682a-4c0f-45e6-a2f1-06b7961c235d	<Binary data>	<Binary data>
04f95c04-d73e-4b11-9180-08f3bc9d3353	<Binary data>	<Binary data>
37ec3a25-96e8-4830-8094-08e4efdb2af3	<Binary data>	<Binary data>
5a1daffa-dc1b-4fc8-0236-09d0ada162e4	<Binary data>	<Binary data>
06d77175-c99f-4f30-b353-0b672095d471	<Binary data>	<Binary data>
0c61cd93-6405-42dd-af90-dc371b27267	<Binary data>	<Binary data>
b2a277e6-2b10-4a22-b777-1019e793bdf1	<Binary data>	<Binary data>
4270a27b-63bf-4861-a3a8-3994627b7ee6	<Binary data>	<Binary data>
0cac0dde-981a-4a95-88c5-110c8749ed91	<Binary data>	<Binary data>
a6419d89-8303-4e7e-b64a-135cd0eb457	<Binary data>	<Binary data>
fe9e28c2-4261-4293-b48a-151f69828d0c	<Binary data>	<Binary data>
2130c6f6-f9a6-4a62-b0d3-15a5c0f41c6	<Binary data>	<Binary data>
123ff52-8503-4700-b85a-1669ccdbd0f	<Binary data>	<Binary data>
7eba52b2-07c0-4a0b-9d47-17c0c381d008	<Binary data>	<Binary data>

Ilustración 28 – Resultados de la Tabla Migracion_Layout

7.2.2 Proyecto PREMIS.Analisis

Con el fin de que el usuario pueda realizar un análisis de la información y del proceso de migración se desarrolló un módulo adicional llamado PREMIS.Analisis (Ver Ilustración 29), este es un proyecto en Asp.Net aplicación de tipo webforms el cual posee dos opciones generales: ver el listado de los documentos migrados y un formulario de tableros el cual permite el diseño por parte del usuario final de Dashboards cuyo origen de datos es la base de datos utilizada para el proceso de migración.

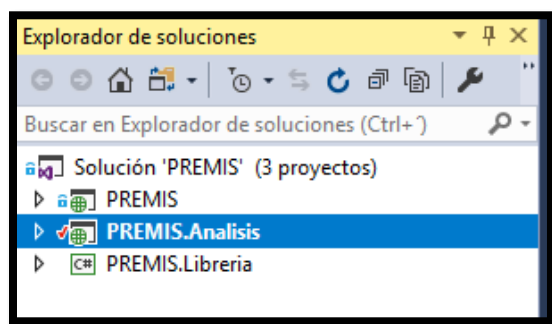


Ilustración 29 – Proyecto PREMIS.Analisis

Al ejecutar este proyecto el usuario verá el formulario principal (Ver Ilustración 30), en el cual están visibles las dos opciones ya mencionadas.

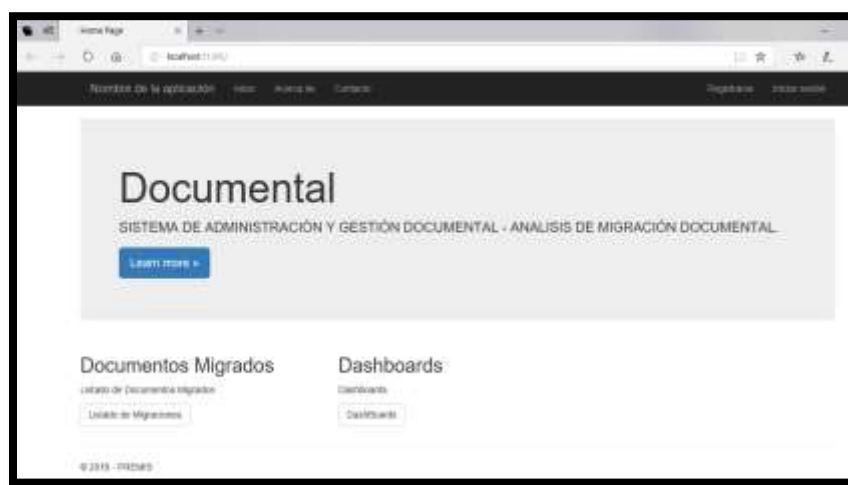


Ilustración 30 – Formulario Principal del Proyecto PREMIS.Analisis

7.2.2.1 Listado de Documentos Migrados

Al dar clic en la opción *Listado de Migraciones* el programa cargará el formulario tal como se muestra en la Ilustración 31:

[illegible]

Ilustración 31 – Formulario listado de documentos migrados

Cada fila que presenta la tabla de este formulario posee la opción de ver el documento ya migrado y la opción para ver el documento PREMIS (Ver Ilustración 32).

Ilustración 32 – Vista de un documento seleccionado en la tabla

7.2.2.2 Diseño de DashBoards

La segunda opción de esta aplicación permite que el usuario cree y consulte Dashboards a partir de la información disponible en la base de datos utilizada para el proceso de Migración Documental.

El formulario que carga la solución cuando el usuario selecciona esta opción se muestra en la Ilustración 33.

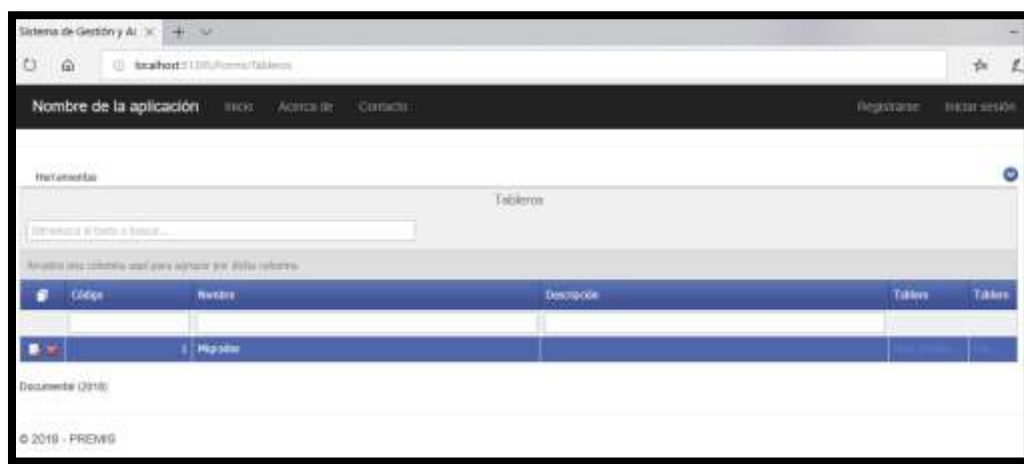


Ilustración 33 – Formulario de Diseño de Dashboards

El formulario carga inicialmente un listado con los tableros ya existentes, dispone también de los botones necesarios para crear, editar y eliminar tableros.

Cuando el usuario selecciona la opción de nuevo el sistema carga un formulario adicional para el diseño del nuevo tablero.

Inicialmente, el usuario debe asignar un nombre al nuevo dashboard (Ver Ilustración 34) y luego dar clic en la opción *Create data source* para crear la conexión a la fuente de datos.

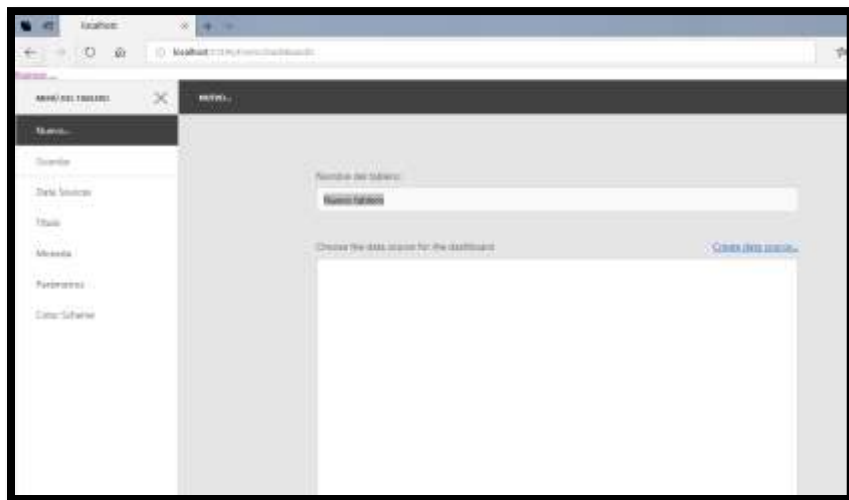


Ilustración 34 –Nuevo Dashboard

Se debe seleccionar la conexión llamada *Origen*, la cual está configurada en el archivo *web.config* del proyecto y que apunta a la base de datos utilizada para el proceso de Migración (Ver Ilustración 35).



Ilustración 35 –Selección origen de datos

Realizado el paso anterior, el usuario deberá crear la consulta deseada a la base de datos, utilizando para esto el asistente que despliega la aplicación (Ver Ilustración 36).



Ilustración 38 –Instrucción SQL generada

Al dar clic en el botón **Terminar**, se abre el formulario de diseño del nuevo dashboard (Ver Ilustración 39):



Ilustración 39 –Formulario para nuevo Dashboard

Luego se da clic en el botón **Crear**

Aparece el formulario de diseño ya configurado con su consulta a un origen de datos.

Se procede a la elaboración dentro del Dashboard (Ver Ilustración 40), de tablas, graficas, entre otros, utilizando la barra de herramientas para insertar todos los controles deseados.

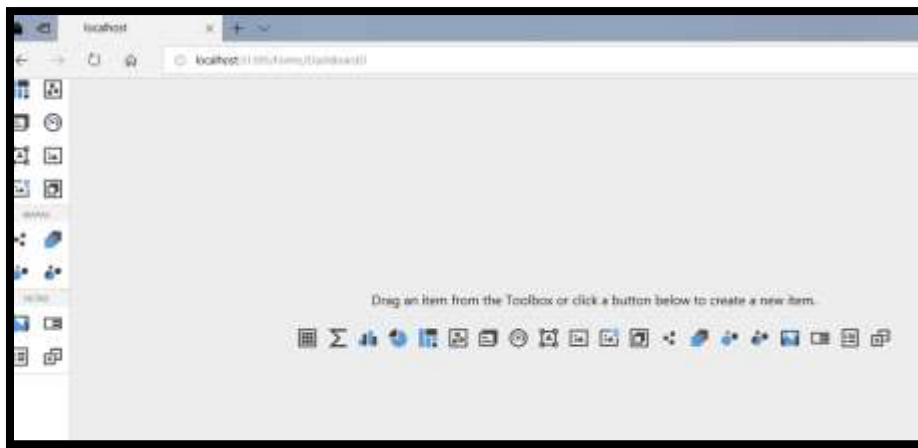


Ilustración 40 – Dashboard y sus propiedades

Por ejemplo, inicialmente le podemos agregar una rejilla (Ver Ilustración 41):

File (DocId)	Path	Filesize (Kb)	File Size (Mb)	Status
121	E:\Masatel\Sistema\Documentos\Documentos\Originals\Documents\121.pdf	2016	2016	True
122	E:\Masatel\Sistema\Documentos\Documentos\Originals\Documents\122.pdf	2016	2016	True
123	E:\Masatel\Sistema\Documentos\Documentos\Originals\Documents\123.pdf	2016	2016	True
124	E:\Masatel\Sistema\Documentos\Documentos\Originals\Documents\124.pdf	2016	2016	True
125	E:\Masatel\Sistema\Documentos\Documentos\Originals\Documents\125.pdf	2016	2016	False
126	E:\Masatel\Sistema\Documentos\Documentos\Originals\Documents\126.pdf	2016	2016	True
127	E:\Masatel\Sistema\Documentos\Documentos\Originals\Documents\127.pdf	2016	2016	True
128	E:\Masatel\Sistema\Documentos\Documentos\Originals\Documents\128.pdf	2016	2016	True
129	E:\Masatel\Sistema\Documentos\Documentos\Originals\Documents\129.pdf	2016	2016	True
130	E:\Masatel\Sistema\Documentos\Documentos\Originals\Documents\130.pdf	2016	2016	False
131	E:\Masatel\Sistema\Documentos\Documentos\Originals\Documents\131.pdf	2016	2016	True
132	E:\Masatel\Sistema\Documentos\Documentos\Originals\Documents\132.pdf	2016	2016	True

Ilustración 41 –Ejemplo Dashboard con una Rejilla

Ahora en la Ilustración 42 vemos como le agregamos un control chart:

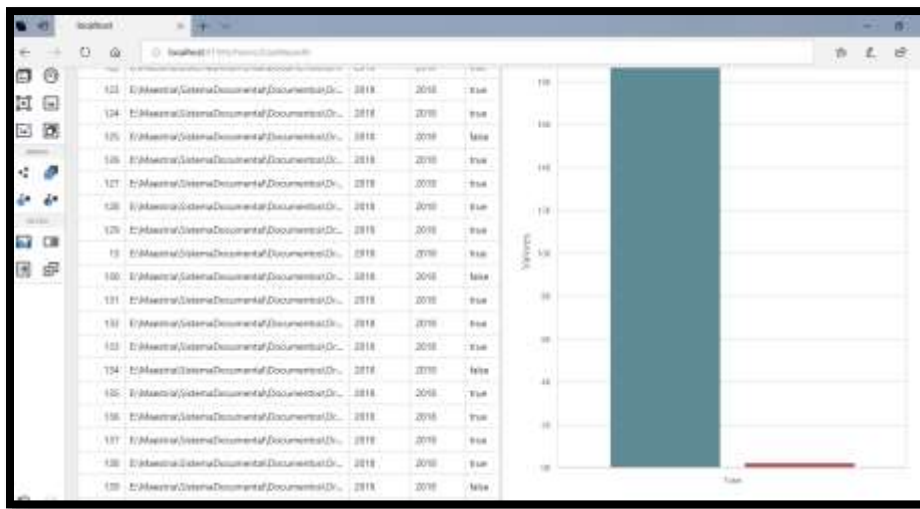


Ilustración 42 –Ejemplo Dashboard con una Grafica

7.2.2.3 Consulta de Dashboards

Existe una opción que permite consultar el Dashboard, tal como se detalla en la Ilustración 43:

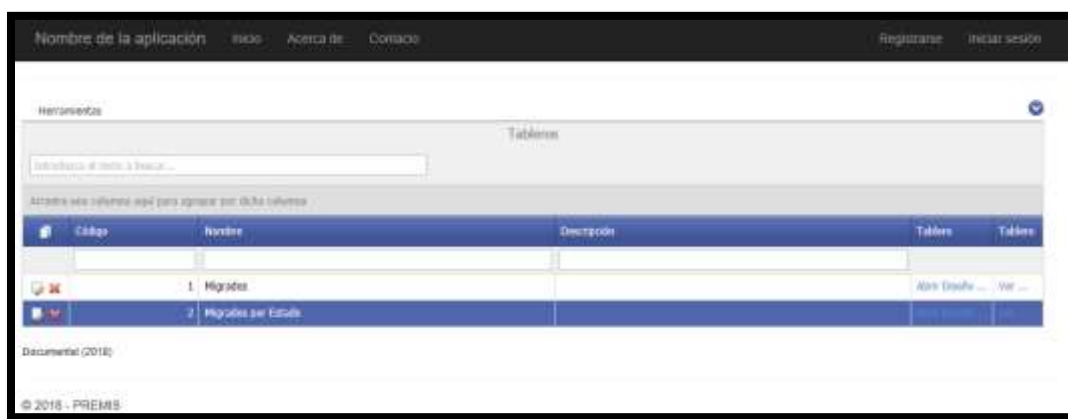


Ilustración 43 –Consulta de un Dashboard

Se selecciona el dashboard deseado y se da clic en la opción ver.

7.3 Resumen de Evidencias y Resultados

Como evidencias del trabajo podemos listar:

- Documentación relacionada con el proyecto
- Documento de Caso de Uso
- Documento de la arquitectura de la Base de Datos
- Documento de la arquitectura del Modelo de Clases
- Programas fuentes de los tres proyectos desarrollados

8. Validación de Resultados con BI

Para el análisis de los resultados de las pruebas del prototipo de migración, se usó la metodología de minería de datos CRISP-DM, y se desarrolló en las siguientes fases:

8.1 Comprensión del problema

Se hizo una revisión del proceso de migración y se determinó como problema: *“La identificación de los patrones y tendencias en los resultados del proceso de migración documental, para la toma de decisiones”*; aunque lo anterior es una hipótesis, solo será resuelta con la implementación de la cada una de las fases de la metodología de CRISP-DM

Si lo anterior es el problema, como casos de éxitos se pueden establecer:

- Identificar los patrones principales
- Mejorar los tiempos de respuesta en la migración

8.2 Comprensión de los datos

En esta fase se identificaron las diferentes fuentes de datos relacionadas al proceso de migración documental y que nos sirven de insumo para aplicar algoritmos de minería que permitan la toma de decisiones

Información identificada:

- Las tablas de las bases de datos donde se registran los resultados del proceso de migración son: LogMigracion y vIndicesDocumentoTexto. Estas tablas están definidas así:

Tabla LogMigracion

Registra la información de los documentos migrados. Comprende los campos id, fila, archivo, hora de inicio, hora fin, estado y errorM.

COLUMNAS	DESCRIPCION
ID	Consecutivo numérico
FILA	Campo consecutivo del archivo de migración (metadatos)
ARCHIVO	Nombre del archivo origen
HORA INICIO	Hora y segundo en que inicio el proceso de migración
HORA FIN	Hora y segundo en que finalizo el proceso de migración
ESTADO	Resultado del proceso de migración, true o false
ERRORM	Mensaje cuando el proceso de migración falla.

Tabla 1 – Campos de la tabla LogMigracion

Tabla vIndicesDocumentoTexto

Almacena resultados de los índices que fueron incorporados en los documentos migrados.

COLUMNAS	DESCRIPCION
ID_INDICEDOCUMENTO	Id del índice del documento
ID_DOCUMENTO	Id del documento
ID_UNIDADDOCUMENTAL	Id de la unidad documental
ID_TIPODOCUMENTAL	Id tipo de unidad documental
ID_INDICE	Id del índice
INDICE	Descripción del índice
VALOR	Valor del índice
S_TIPO	Tipo
ID_INDICEUNIDADDOC	Índice Unidad Documental
B_OBLIGATORIO	Obligatorio
N_ORDEN	Orden
B_ACTIVO	Activo

S_MASCARA	Mascara
N_LONGITUD	Cantidad de caracteres
S_AYUDA	Ayuda

Tabla 2 – Campos de la tabla vIndicesDocumentosTexto

Archivo de migración con la definición de los metadatos que se cargaron a los documentos.

COLUMNAS	DESCRIPCION
ID	Consecutivo numérico
RUTA	Ruta de ubicación origen del documento
ENTIDAD	Entidad Controladora
SERIE	Serie a la que perteneces el documento
SUBSERIE	Subserie asociada a la serie del documento
UNIDAD DOCUMENTAL	Unidad Documental a la que pertenece la subserie del documento
INDICE 1	Índice asociado a la unidad documental
VALOR INDICE 1	Valor del índice
INDICE 2	Índice asociado a la unidad documental
VALOR INDICE 2	Valor del índice
INDICE 3	Índice asociado a la unidad documental
VALOR INDICE 3	Valor del índice
INDICE 4	Índice asociado a la unidad documental
VALOR INDICE 4	Valor del índice
INDICE 5	Índice asociado a la unidad documental
VALOR INDICE 5	Valor del índice
INDICE 5	Índice asociado a la unidad documental
VALOR INDICE 5	Valor del índice

Tabla 3 – Campos del archivo migración (metadatos)

8.3 Preparación de los datos

Se hace la recolección de los datos de las fuentes identificadas. La data correspondiente a la base de datos se consulta y se exporta en formato .csv (delimitado por tabulaciones). La data de los metadatos de los documentos se tiene en archivo .txt.

Para el análisis de los datos, se hace un perfilamiento de los datos con el fin de conocer su estructura, contenido, formato y nivel de calidad, para esto utilizaremos la herramienta DQ Analyzer (Benefits & Advantages, 2011)

8.3.1 Perfilamiento Tabla Resultados de Migración

Se hace el análisis a cada campo de la tabla para obtener los resultados con respecto a valores null, valores únicos, tipo de campo, estadísticas, tamaño.

En la aplicación se selecciona la opción de nuevo perfil, se elige el origen de los datos donde tenemos el archivo de la tabla **Log Migración** y se proceder a ejecutar; la herramienta arroja los siguientes resultados que se detallan en la Ilustración 44 :

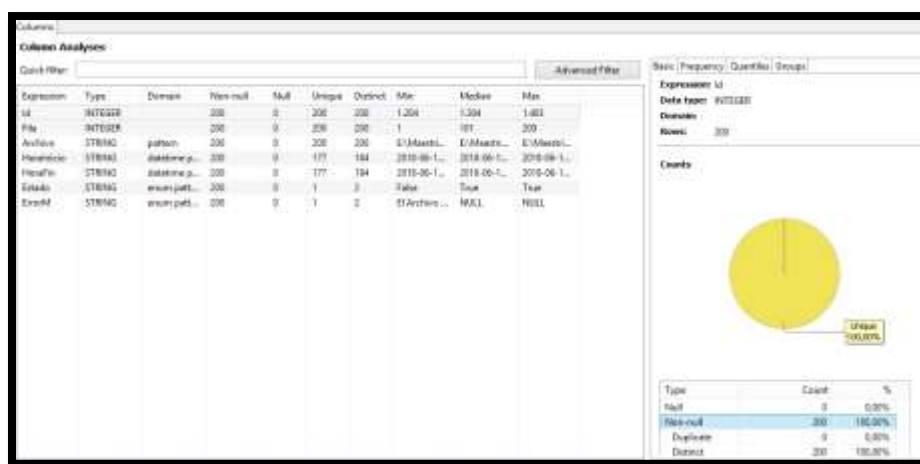


Ilustración 44 –Perfil DQ Analyzer Tabla Log Migración

Se analizaron 200 registros de esta tabla, obteniéndose la siguiente interpretación para cada campo:

Id: Campo tipo integer, ningún valor null, todos son valores únicos

Fila: Campo tipo integer, ningún valor null, todos son valores únicos, son registros numéricos consecutivos.

Archivo: Campo tipo string, ningún valor null, todos son valores únicos, cantidad de registros 200, longitud de 68 caracteres.

HoraInicio: Campo tipo string, no posee valores null, valores distintos 184, longitud de 23 caracteres

HoraFin: Campo tipo string, no posee valores null, valores distintos 184, longitud de 23 caracteres

Estado: Campo tipo string, no posee valores null, valores distintos 2, corresponde a true y false, longitud de caracteres 4.

ErrorM: Campo tipo string, no posee valores null, valores distintos 2, longitud de caracteres 94

8.3.2 Perfilamiento Tabla Valores Índices Migrados

Se hace el mismo análisis para el perfilamiento de los datos realizado con la tabla de log de migración pero esta vez a la data exportada de la tabla *vIndicesDocumentoTexto*

Se ejecuta la herramienta de DQ Analyzer, ver Ilustración 45; que arroja los siguientes resultados en el explorer de la aplicación:

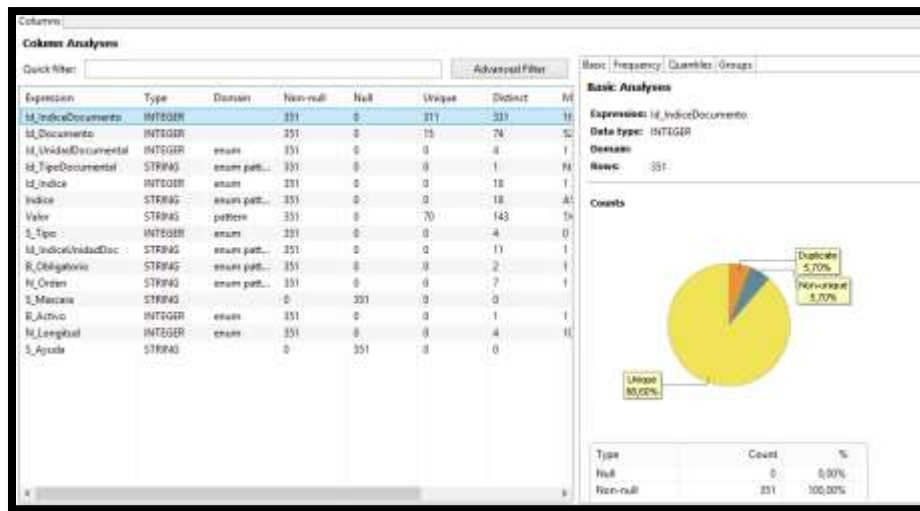


Ilustración 45 –Perfil DQ Analyzer Tabla vIndicesDocumentoTexto

Se analizaron 351 registros de esta tabla, obteniéndose la siguiente interpretación para cada campo:

- **Id_IndiceDocumento:** campo tipo integer, no tiene valores null, tiene 331 valores distintos
- **Id_Documento:** campo tipo integer, no tiene valores null, tiene 74 valores distintos
- **Id_UnidadDocumental:** campo tipo integer, no tiene valores null, tiene 4 valores distintos
- **Id_TipoDocumental:** campo tipo string, no tiene valores null, tiene 1 solo tipo de valor, la longitud del campo es 4 caracteres
- **Id_Indice:** campo tipo integer, no tiene valores null, tiene 18 valores distintos
- **Índice:** campo tipo string, no tiene valores null, tiene 18 valores distintos, longitud del campo 19 carácter
- **Valor:** campo tipo string, no tiene valores null, valores distintos 143, longitud del campo 31 caracteres
- **S_Tipo:** campo tipo integer, no tiene valores null, valores distintos 4
- **Id_IndiceUnidadDocumental:** campo tipo string, no tiene valores null, valores distintos 11, longitud del campo de 4 caracteres

- **B_Obligatorio:** campo tipo string, no tiene valor null, valores distintos 2, longitud del campo de 4 caracteres
- **N_Orden:** campo tipo string, no tiene valor null, valores distintos 7, longitud del campo de 4 caracteres
- **S_Mascara:** campo tipo string, todo está null
- **B_Activo:** campo tipo integer, no tiene valores null, valores distintos 1
- **N_Longitud:** campo tipo integer, no tiene valores null, valores distintos 4
- **S_Ayuda:** campo tipo string, todo está null

Se concluye, después de haber aplicado las técnicas de análisis que la calidad de los datos es alta, algunos valores nulos que no afecto la modelación, valores duplicados de campos que lo permiten, por lo que no habría necesidad de hacer un tratamiento especial. Adicional el proceso de migración cuenta con las validaciones que garantizan que la información errada no sea almacenada en las tablas.

8.4 Modelado

Se aplicaron los algoritmos de minería de datos para la modelación de los datos preparados en la fase anterior de preparación de los datos.

Para aplicar estos algoritmos de minería de datos se utilizó la herramienta de minería WEKA (Waikato Environment for Knowledge Analysis). Weka es un software de código libre que tiene una colección de algoritmos para tareas de minería de datos. Los algoritmos se pueden aplicar a un conjunto de datos o desde su propio código java. Contiene herramientas para el procesamiento de datos entre las que se encuentran: clasificación, regresión, clustering, reglas de asociación y visualización.(Frank, Hall, & Witten, 2016)

8.4.1 Modelo con Weka de los Resultados de la Migración

Se usa la herramienta Weka para el procesamiento de los datos; en la pestaña *Preprocess* se selecciona Open File, se busca el archivo que para este caso corresponde a los datos con los Resultados de la Migración y se procede a procesarlo.

Ya en la vista explorer podemos visualizar y aplicar distintos algoritmos para el conjunto de datos y aplicar los filtros que queramos a los atributos (Ver Ilustración 46). Estos algoritmos son:

- Classify: Algoritmos de clasificación y regresión
- Cluster: Agrupación
- Associate: Asociación

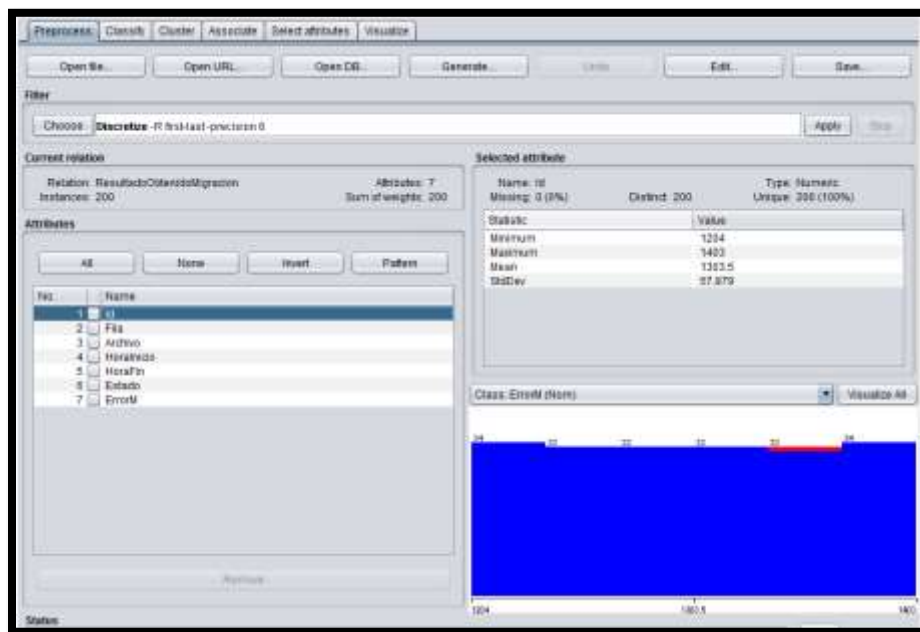


Ilustración 46 –Modelo con Weka de los Resultados de la Migración

Para la técnica de clasificación se seleccionó el algoritmo *Tree* (árboles)/J48 por ser un ágil y preciso (Ver Ilustración 47). En *test options* se seleccionó el tipo *Training set* para analizar todos los registros y como atributo se filtró por Estado; obteniéndose los siguientes resultados:

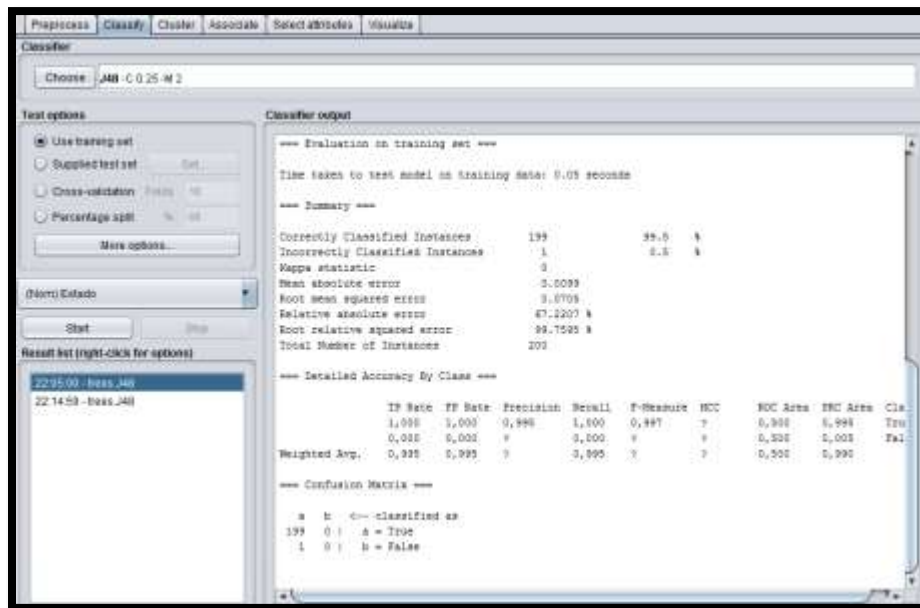


Ilustración 47 –Técnica de Clasificación

Se observa una muestra de los registros, que corresponden a 200 documentos empleados para la validación del prototipo de migración. Los resultados del prototipo arrojaron un solo documento no migrado, por lo que este insumo de entrada al modelo de minería hace que no se presenta mucha variabilidad para la interpretación.

Del modelo se observa que la matriz de confusión es correcta porque su diagonal obtuvo los mayores valores. El árbol solo tiene una hoja (Ver Ilustración 48), y el porcentaje de error fue de 0.5%. Lo que nos quiere decir que los resultados del proceso de migración son satisfactorios y que tiene una muy baja probabilidad de fallar (de no migrar el documento).



Ilustración 48 –Árbol de decisión

Para la técnica de clúster, se escogió el algoritmo *SimpleKMeans*, la opción de *Training set* para que tenga en cuenta todos los atributos y se inicia el proceso. Como resultado se obtuvo, (Ver Ilustración 49):

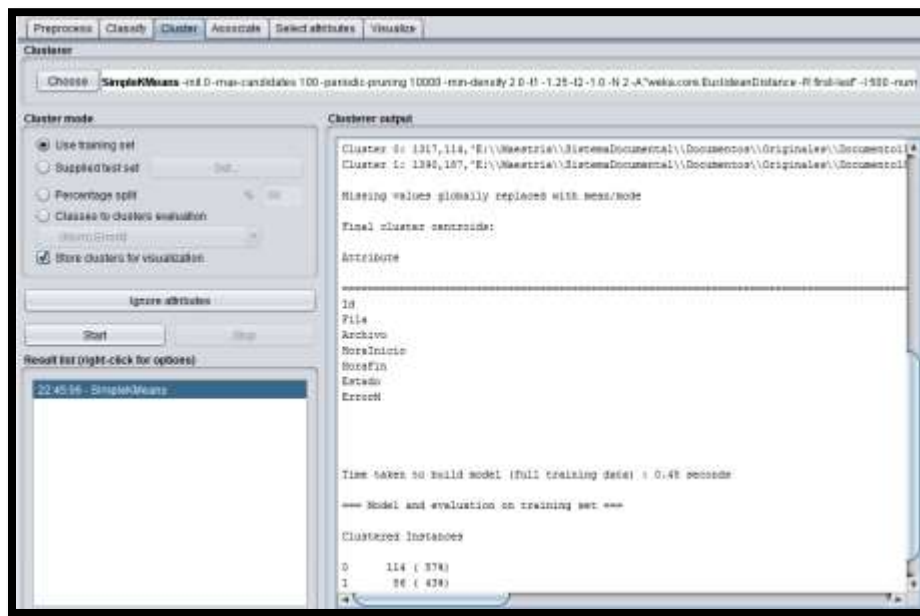


Ilustración 49 –Técnica de Clúster

Se puede observar que clúster hizo 8 iteraciones y particionó los datos en dos grupos para hacer unos análisis independientes. El clúster #0 tiene 114 registros y el #2 tiene 187.

Analizando la matriz de confusión me dice indica el porcentaje de registros en cada clúster, siendo para el 1 el 57% y el 2 el 43%.

Para la técnica de asociación se seleccionó el algoritmo *A priori* pero el botón de Start no se activó por lo que este algoritmo no aplicaría para nuestra data (Ver Ilustración 50):

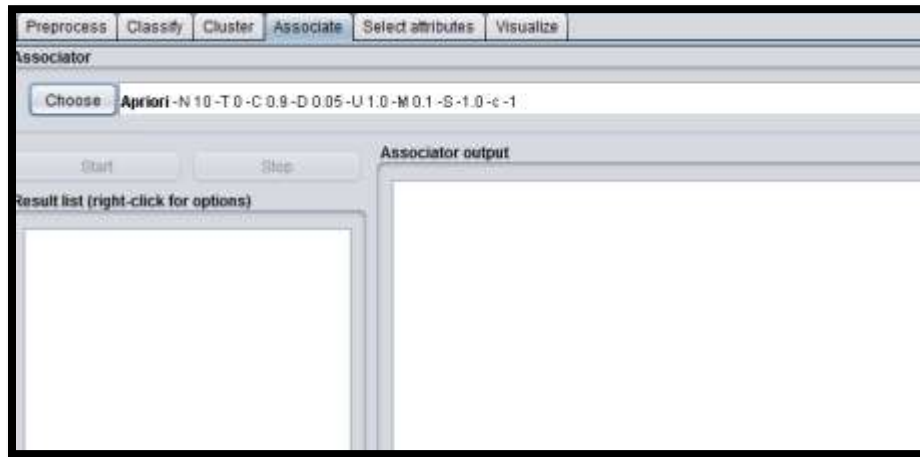


Ilustración 50 –Técnica de Clasificación

8.4.2 Modelo con Weka de los Valores Índices Migrados

Se modela los datos correspondientes a los valores índices migrados y se obtiene los siguientes resultados en el explorador de Weka, Ver Ilustración 51:

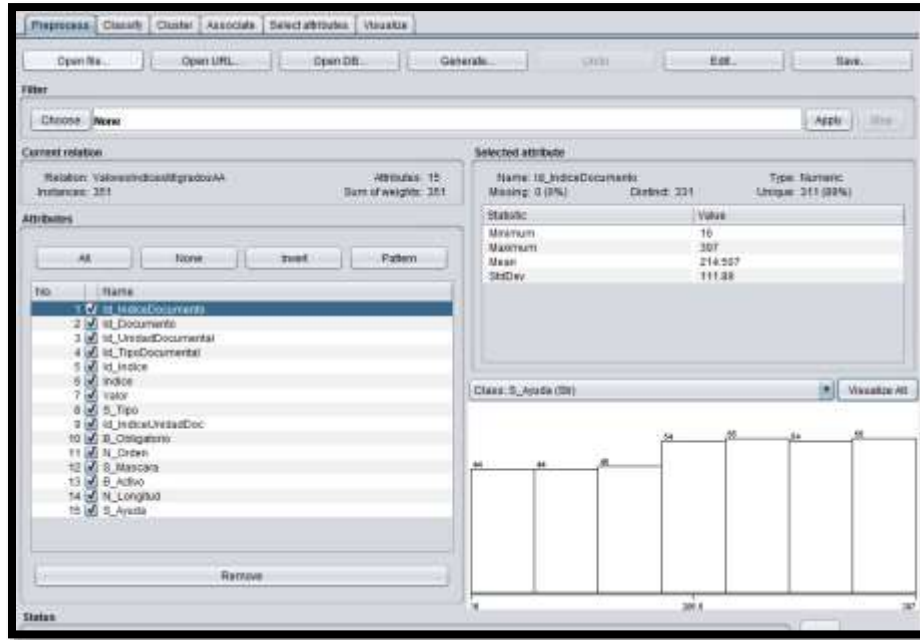


Ilustración 51 –Modelo con Weka de los Valores Índices Migrados

En la Ilustración 52 se observa el explorador donde se pueden visualizar uno a uno los atributos o todos a la vez:

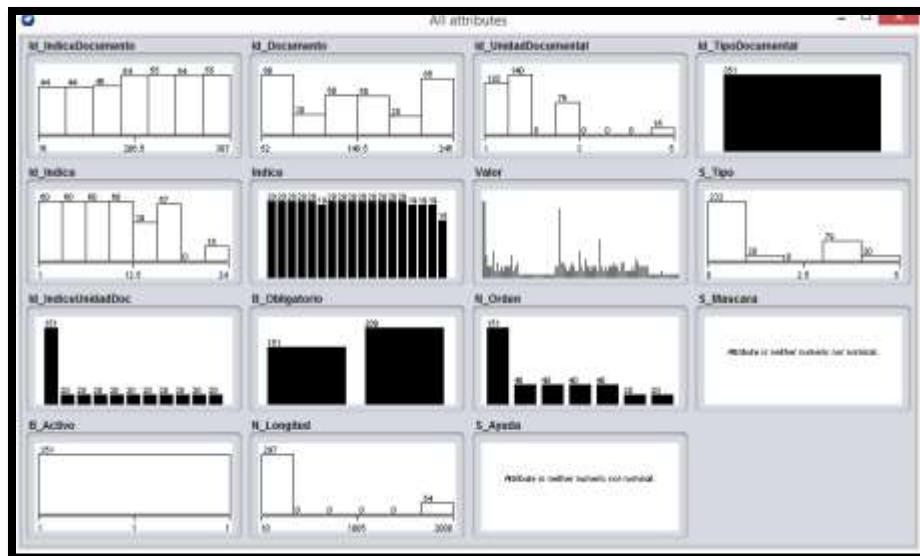


Ilustración 52 –Visualización de los atributos de los Valores Índices Migrados

Para estos datos no se pudieron modelar con ninguno de los algoritmos (clasificación, clúster y asociación) porque es una data que no hay correlación para análisis entre ellas, hay valores null y otros son campos de id, porque corresponden con los índices. Por lo que de esta tabla no se puede interpretar ningún resultado para toma de decisión.

8.5 Evaluación e Implementación

Se evalúa los resultados obtenidos en los modelos y se interpreta lo siguiente: solo se pudo aplicar algoritmo de minería de datos a la data de los resultados de la migración. No se presentó muchas variaciones entre las diferentes técnicas, porque el campo clave que es el estado no tuvo sino un valor diferente que correspondía a un false (1 solo documento que no se migro). Con estos datos no se puede tomar patrones para predicciones, lo que se puede interpretar es que el prototipo de migración no falla en si por cómo está definido, sino por errores en la formulación del archivo txt de la metada o por ausencia del archivo origen al que se está invocando. Como resultado no hay caso de éxito porque el problema inicial no se presentó.

Conclusiones

1. Con el diseño del sistema de información propuesto en este trabajo, se puede realizar la programación de un sistema de información enfocado en cubrir todas las necesidades de un Centro de Administración Documental (CAD), tanto para empresas públicas y privadas; porque cumple con la normatividad vigente establecida por el Archivo General de la Nación.
2. Con la documentación actual de los requerimientos funcionales, no funcionales y arquitectónicos, cualquier persona o empresa está en la capacidad de desarrollar el sistema de gestión documental y hacer sus ajustes a la medida que se requieran; su diseño está orientado a la programación orientada a objetos y su base de datos está debidamente normalizada y bien documentada en su diccionario.
3. Al obtener una efectividad casi del 100% en el proceso de migración; se concluye que el prototipo migra cualquier documento que se encuentre en formato libre pdf y garantiza su disponibilidad en el tiempo; lo anterior por la ingesta de sus metadatos.
4. El módulo de migración documental es una valiosa herramienta, que permite incorporar a un conjunto de documentos PDF, un archivo en formato XML con la información del estándar PREMIS y subir estos documentos a la base de datos propuesta para el sistema de gestión y administración documental. Basados en los programas fuentes y en el diseño del módulo de migración, los desarrolladores y entidades interesadas pueden evolucionar este prototipo a un sistema de migración mucho más complejo y automático, que vaya incorporando a través del tiempo las nuevas tendencias y/o estándares en materia de migración documental y preservación de documentos digitales y que les ahorre tiempo y dinero al enfrentarse a este tipo de procesos.

5. El prototipo de migración puede ser utilizado por la industria del software en entidades que desarrollen productos a la medida. El funcionamiento del módulo está soportado en la información de los metadatos PREMIS, lo que lo hace independiente de la arquitectura y del diseño de la base de datos del sistema.
6. Se deben almacenar más información del proceso de migración. Actualmente las tablas no permiten predecir patrones de relación entre el volumen con el tiempo que dura en migrarse el documento.
7. Con la documentación técnica completa, se puede avanzar directamente a la fase de programación, disminuyendo con esto altamente los costos asociados con la implementación del sistema. Estos diseños y su documentación pueden ser utilizados y adaptados a necesidades puntuales de cada entidad.

Recomendaciones

- Seguimiento a tiempos de respuesta con pruebas reales al prototipo de migración
- Almacenar más información en el proceso de migración para futuros análisis

Temas propuestos como continuación de este trabajo

- Desarrollar e implementar el diseño del sistema de gestión documental.
- Automatizar el proceso de la creación del archivo previo de los metadatos.
- Proponer nuevas alternativas en la búsqueda de garantizar la preservación documental.
- Desarrollar investigaciones y desarrollos enfocados a mejorar la seguridad de los documentos y a garantizar su preservación a largo plazo.

Referencias bibliográficas

- Agudelo, M. (2009). Los Metadatos. *Gestión Contenidos Educ. Virtual Calid.*, 5. Retrieved from http://aprendeenlinea.udea.edu.co/lms/men/docsoac3/0301_metadatos.pdf
- Archivo General de la Nación. (2001). Mini Manual, Tablas de retención y transferencias documentales. Bogotá, D.C.: Archivo General de la Nación.
- Archivo General de la Nación. (2012). Guía del Uso del PDF/A, 1–30.
- Archivo General de la Nación. (2014). Guía para la formulación de un esquema de metadatos para la gestión de documentos, 1–26.
- Archivo General de la Nación. (2017). <http://www.archivogeneral.gov.co/mision-vision>.
- Awaad, M. H., Krauss, H., & Schmatz, H. D. (1978). *Advanced Praise for The Unified Modeling Language Reference Manual, Second Edition. Zentralblatt für Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene. Erste Abteilung Originale. Reihe A: Medizinische Mikrobiologie und Parasitologie* (Vol. 240).
- Benefits, B., & Advantages, T. (2011). Ataccama DQ Analyzer Product Sheet.
- Biblioteca Nacional de España. (2015). DICCIONARIO DE DATOS PREMIS De Metadatos de Preservación, *Versión 2.*, 244.
- Camarena Sagredo, M., Gamaliel, J., Espinosa, T., Reyes, M., García, L., & de Lourdes, M. (2012). Automatización de la codificación del patrón modelo vista controlador (MVC) en proyectos orientados a la Web, *1*, 13.
- Caplan, P. (2009). Entender PREMIS. *Library*, 3–30.
- Cesar Tarazona. (2002). Amenazas Informaticas y Seguridad de la Informacion. *Etek Internacional*, 137–146.
- Chornet, V. G. (2014). Criterios ISO para la preservación digital de los documentos de archivo. *Códices*, *10*(2), 135–150.

- Comité Técnico de Normalización 50 AENOR. (2008). Información y documentación- Procesos de gestión de documentos-Metadatos para la gestión de documentos. Parte 1: Principios. ISO 23081-1: 2006. *Revista Española de Documentación Científica*, 31(2), 273–301. Retrieved from http://www.bibliologia.info/archivos/UNE_ISO_23081_2_Procesos_de_gestion_de_documento_Metadatos_parte_1.pdf
- Consejo de la UE. (2002). Conservar la memoria del mañana, Conservar los contenidos digitales para las generaciones futuras, 25–27.
- Cruz Mundet, J. R., & Díez Carrera, C. (2016). Sistema de Información de Archivo Abierto (OAIS): luces y sombras de un modelo de referencia. *Investigacion Bibliotecologica*, 30(70), 221–247. <https://doi.org/10.1016/j.ibbai.2016.10.010>
- Edin Gutiérrez. (2017). Como hacer una Tabla de Retención Documental. Retrieved from <https://www.nosonpapeles.com/hacer-tabla-de-retencion-documental/>
- ESA. (1994). ESA Software Engineering Standards. *ESA Publications Division, ESA PSS-05(2)*. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:ESA+software+engineering+standards#1>
- Frank, E., Hall, M. A., & Witten, I. H. (2016). The WEKA workbench. *Morgan Kaufmann, Fourth Edition*, 553–571. <https://doi.org/10.1016/B978-0-12-804291-5.00024-6>
- Gallardo, J. A. (2009). Metodología para el desarrollo de proyectos en Minería de Datos CRISP-DM.
- Giusti, M. R. De. (2015). Preservación Digital : metadatos y preservación. *Universidad Nacional de La Plata*, 69.
- ICONTEC. (2005). Iso-19005 pdf/a, 4, 1–3.
- ICONTEC. (2013). GUÍA TÉCNICA COLOMBIANA GTC-ISO-TR 18492, (571).
- ISO. (2015). Norma ISO 14721, 6–9.

- Lavoie, B. (2014). Information System (OAIS) Reference Model : Introductory Guide. *Technology Watch Reports*, 37. <https://doi.org/http://dx.doi.org/10.7207/twr14-02>
- Melrose, J., Perroy, R., & Careas, S. (2015). Carta para la preservación del Patrimonio Digital. *Statewide Agricultural Land Use Baseline 2015*, 1, 1–5. <https://doi.org/10.1017/CBO9781107415324.004>
- Microsoft. (2009). Información General sobre ASP.NET MVC. Retrieved from <https://docs.microsoft.com/es-es/aspnet/mvc/overview/older-versions-1/overview/asp-net-mvc-overview>
- MINTIC. (2011). Sistemas de Gestión de Documentos Electrónicos Guía Nro 6.
- MINTIC. (2017). G . INF . 08 Guía para la gestión de documentos y expedientes electrónicos Guía técnica, 131.
- Miñambres, Ó. P. (2007). Minería de Datos, 8.
- Moine, J. Mi., Haedo, A., & Gordillo, S. (2011). Estudio comparativo de metodologías para minería de datos. *XIII Workshop de Investigadores En Ciencias de La Computación*, 278–281. Retrieved from <http://sedici.unlp.edu.ar/handle/10915/20034>
- Montequín, R., Teresa, M., Cabal, Á., Valeriano, J., Fernández, M., Manuel, J., & Valdés, G. (2003). Metodologías Para La Realización De Proyectos De Data Mining. *VII Congreso Internacional de Ingeniería de Proyectos*, 257–265. Retrieved from http://www.aepro.com/files/congresos/2003pamplona/ciip03_0257_0265.2134.pdf
- Natividad, G. S., & Salamanca, U. De. (2015). PRESERVACIÓN Y CONSERVACIÓN DE DOCUMENTOS DIGITALES, 25.
- Olga Pons, Nicolás Marín, Juan Miguel Medina, Silvia Acid & Vila, M. A. (2005). Introducción a las Bases de Datos: El modelo relacional.
- Reynaldo Antonio Espinosa, G. B. A. R. INSTRUCTIVO PARA VALORACIÓN Y DISPOSICIÓN FINAL DE DOCUMENTOS (2014).
- Rodríguez A, C. M., & Gaitán Gómez, H. (2011). *Sistemas de Gestión Documental - Guía*

práctica para el diseño e implementación. (E. Bonventuriana, Ed.). Buenaventura.

SEDIC. (2003). Dublin Core (DCMI). ISO 15836-2003. Retrieved from

<https://www.sedic.es/autoformacion/metadatos/tema7.htm>

Sinha, R. (2010). Prácticas Ágiles - Desarrollo de software con un enfoque ágil, 1–6.

Trigas Gallego, M., & Domingo Troncho, A. C. (2012). Gestión de Proyectos Informáticos.

Metodología Scrum. *Openaccess.Uoc.Edu*, 56. Retrieved from

<http://www.quimbiotec.gob.ve/sistem/auditoria/pdf/ciudadano/mtrigasTFC0612memoria.pdf%5Cnhttp://openaccess.uoc.edu/webapps/o2/bitstream/10609/17885/1/mtrigasTFC0612memoria.pdf>

UNESCO. (2003). Directrices Para La Preservación Del Patrimonio, 186.

UNESCO. (2009). ¿Qué es la UNESCO?¿Qué hace?

Anexos

- ANEXO 1 - CASOS DE USO
- ANEXO 2 - MODELO DE CLASES
- ANEXO 3 - MODELO FÍSICO DE DATOS
- ANEXO 4 - MODELO DE CLASES MÓDULO DE MIGRACIÓN PREMIS
- ANEXO 5 - SCRIPT PARA LA GENERACIÓN DE LA BASE DE DATOS DE MIGRACIÓN
- ANEXO 6 – MANUAL DE USUARIO DEL PROTOTIPO DE MIGRACION