

Evolución de la Ciencia de Datos y Big Data en la Gestión Documental

Wagner Alonso Fernández Villalobos

Asesora

María Alejandra Varona Taborda

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI

Especialización en Ciencia de Datos y Analítica

2025

Nota de Aceptación

Nombre Director de Trabajo de Grado

Jurado

Jurado

Dedicatoria

Dedico esta monografía a mis familiares, en especial a mi madre, que siempre me ha dicho “*vamo’ pa’ lante*” y a mi esposa con su frase heredada “*si se puede*”, quienes siempre me incentivaron a pensar críticamente, para buscar respuestas a preguntas; también a todas esas personas que siempre dudaron que llegaría lejos y a todos los que hacemos el trabajo invaluable de ser archivistas, los cuales somos guardianes de la información, para servir constantemente a la sociedad.

Agradecimiento

Agradezco cordialmente a la profesora María Alejandra Varona Taborda, por su invaluable guía y apoyo durante la realización de este proyecto de grado, sus conocimientos y orientación fueron fundamentales para el desarrollo exitoso de esta monografía.

Resumen

Esta monografía tiene como propósito analizar la evolución e impacto de la Ciencia de Datos y Big Data en la Gestión Documental, para ello, se investiga cómo la gestión documental ha transitado desde métodos tradicionales hacia prácticas optimizadas por tecnologías digitales, se examinan los conceptos fundamentales de la Ciencia de Datos y Big Data, destacando las principales técnicas y herramientas empleadas en este ámbito. Además, se analizan casos de estudio en organizaciones que han implementado estas tecnologías con éxito, evidenciando los beneficios obtenidos en términos de acceso, seguridad y eficiencia en los procesos documentales; finalmente, se evalúan los retos y oportunidades que estas herramientas representan, destacando su importancia en la toma de decisiones informadas. Esta recopilación aporta una visión integral de cómo la Ciencia de Datos y Big Data pueden transformar la gestión documental en empresas u organizaciones, optimizando procesos y adaptándose a las demandas del entorno digital actual.

Palabras clave: Big Data, Ciencia de Datos, Conocimiento, Gestión Documental, Metadatos.

Abstract

The purpose of this monograph is to analyze the evolution and impact of Data Science and Big Data on Document Management, for this, it investigates how document management has moved from traditional methods to practices optimized by digital technologies, examines the fundamental concepts of Data Science and Big Data, highlighting the main techniques and tools used in this field. In addition, case studies in organizations that have successfully implemented these technologies are analyzed, evidencing the benefits obtained in terms of access, security and efficiency in document processes; Finally, the challenges and opportunities that these tools represent are evaluated, highlighting their importance in making informed decisions. This compilation provides a comprehensive vision of how Data Science and Big Data can transform document management in companies or organizations, optimizing processes and adapting to the demands of today's digital environment.

Keywords: Big Data, Data Science, Document Management, Knowledge, Metadata.

Tabla de Contenido

Introducción	11
Justificación	12
Objetivos.....	14
Objetivo General	14
Objetivos Específicos.....	14
Planteamiento y Descripción del Problema	15
Marco de Referencia	16
Antecedentes	16
Ciencia de Datos	17
Big Data	17
Gestión Documental.....	18
Marco Normativo.....	18
Marco Normativo Internacional.....	18
Marco Normativo en Colombia	19
Implicaciones y Usos Prácticos	20
Mejora de Procesos en Gestión Documental	20
Toma de Decisiones Basada en Datos.....	20
Evolución de la Gestión Documental en la Era Digital: Transformaciones Impulsadas por la Ciencia de Datos y Big Data.....	21
Década 1960 – 1970: Fundamentos y Comienzos de la Digitalización.....	21
Década de 1980: Era de las “Personal Computers” e Institucionalización de los Archivos.....	23
Década 1990 – 2000: Internet y Globalización.....	26

Década de 2010: Computación en la Nube, Digitalización y Ética en la Gestión Empresarial	28
Década de 2020: Inteligencia Artificial y los Datos como su Combustible.....	30
Herramientas y Técnicas de Ciencia de Datos y Big Data en la Gestión Documental: Impacto en la Eficiencia y Optimización de Procesos.....	33
Beneficios, Desafíos y Casos de Estudio en la Implementación de Ciencia de Datos y Big Data en la Gestión Documental.....	37
Empresa LAIEX.....	37
Policía Nacional de Colombia.....	39
Google	44
Conclusiones	47
Recomendaciones	48
Referencias Bibliográficas	49

Lista de Tablas

Tabla 1 <i>Uso de Herramientas y Tecnologías para el Proceso de Gestión Documental.</i>	35
--	----

Lista de Figuras

Figura 1 <i>Organización y Digitalización de Archivos.</i>	33
Figura 2 <i>Acceso al Gestor de Documentos Policiales - GEPOL.</i>	41
Figura 3 <i>Tablero Comportamiento Documentos</i>	43
Figura 4 <i>Captura de Pantalla Aplicación Google Drive</i>	45
Figura 5 <i>Uso Heramienta Document AI</i>	46

Introducción

En un mundo cada vez más digitalizado, la gestión de la información se ha convertido en un pilar fundamental para las organizaciones, la elaboración de datos estructurados y no estructurados, ha superado con creces las capacidades de los sistemas tradicionales de gestión documental; bajo este contexto, la ciencia de datos y el big data han emergido como herramientas poderosas para transformar la forma en que las empresas elaboran, recolectan, almacenan, analizan y utilizan su patrimonio documental. En este entregable, se expone la evolución de este trinomio (*Ciencia de Datos, Big Data y Gestión Documental*), desde los primeros acercamientos hasta las aplicaciones más actuales y cómo estas tecnologías han revolucionado el proceso de gestión documental, permitiendo extraer valor de los datos, prestar un servicio eficiente y tomar decisiones más informadas.

Justificación

En la era digital actual, el crecimiento exponencial de los datos plantea tanto oportunidades como desafíos para diversos sectores, incluyendo la Gestión Documental. La gran cantidad de información, conocida como Big Data, requiere herramientas y enfoques avanzados que permitan procesar, analizar y aprovechar estos datos de manera eficiente. En este contexto, la Ciencia de Datos emerge como una disciplina clave, al ofrecer técnicas y metodologías para extraer conocimiento significativo de grandes volúmenes de datos, mejorando así los procesos de clasificación, conservación y acceso a documentos en formatos físicos y digitales.

La gestión eficiente de documentos no solo impacta directamente la operación interna de organizaciones públicas y privadas, sino que también juega un papel crucial en el acceso a la información, la toma de decisiones y el cumplimiento de normativas internacionales. En un contexto global, la transformación digital ha llevado a gobiernos y empresas a buscar soluciones innovadoras que optimicen estos procesos, lo que refuerza la pertinencia de estudiar cómo la Ciencia de Datos y el Big Data están moldeando esta área.

Realizar esta monografía es relevante porque aborda un problema fundamental: ¿cómo aprovechar las tecnologías de análisis de datos para hacer frente a los desafíos de la Gestión Documental en la era digital? Este análisis no solo contribuirá a identificar técnicas y herramientas útiles para el sector, sino que también destacará los beneficios y desafíos asociados, como la seguridad de la información, el acceso equitativo y la eficiencia operativa.

A nivel social, la adecuada gestión de documentos es esencial para garantizar derechos fundamentales, como el acceso a la información pública y la transparencia. Además, este estudio podría ser una guía valiosa para organizaciones interesadas en implementar soluciones tecnológicas basadas en Ciencia de Datos, fortaleciendo la capacidad de respuesta ante las

demandas de un entorno cada vez más dinámico y competitivo.

En el ámbito académico y profesional, la investigación contribuirá al campo del conocimiento al sistematizar las principales técnicas y herramientas de Ciencia de Datos aplicadas a la Gestión Documental, sentando las bases para futuros estudios y desarrollos en esta área. Así mismo, este trabajo subraya la importancia de adoptar enfoques basados en evidencia, destacando cómo la analítica de datos puede transformar procesos críticos para el desarrollo sostenible de las organizaciones y las comunidades en general.

Objetivos

Objetivo General

Analizar la evolución de la Ciencia de Datos y Big Data, así como su influencia en la Gestión Documental a lo largo del tiempo.

Objetivos Específicos

Investigar la evolución de la gestión documental con la llegada de la era digital, identificando las principales transformaciones impulsadas por la Ciencia de Datos y Big Data.

Analizar las herramientas y técnicas de Ciencia de Datos y Big Data aplicadas en la gestión documental, destacando su impacto en la eficiencia y optimización de procesos.

Evaluar los beneficios, desafíos y casos de estudio de organizaciones que han implementado estrategias basadas en Ciencia de Datos y Big Data para mejorar el acceso, la seguridad y la toma de decisiones en la gestión documental.

Planteamiento y Descripción del Problema

En la actualidad, la gestión documental enfrenta retos significativos derivados del crecimiento exponencial de los datos, tanto en formatos físicos como digitales, históricamente, los métodos tradicionales de archivo y organización han sido insuficientes para responder a las necesidades de eficiencia, acceso y conservación de la información; con la llegada de herramientas y tecnologías como la Ciencia de Datos y el Big Data, han surgido nuevas oportunidades para transformar estos procesos, pero también nuevos desafíos relacionados con la integración tecnológica, la capacitación de talento humano y el cumplimiento normativo.

El principal problema radica en la capacidad limitada de los sistemas tradicionales de gestión para manejar volúmenes masivos de datos, lo que resulta en pérdidas de información, procesos ineficientes y dificultades en la toma de decisiones informadas. La falta de estándares en el manejo de datos estructurados, semiestructurados y no estructurados, junto con la escasa adopción de herramientas avanzadas de análisis, ha suscitado brechas entre la producción de información y su aprovechamiento estratégico en diferentes sectores.

En este contexto, la evolución de la Ciencia de Datos y el Big Data ofrece un marco prometedor para optimizar el proceso de Gestión Documental. Sin embargo, su implementación requiere superar desafíos técnicos, éticos y culturales que condicionan su efectividad, esta monografía, busca abordar estas problemáticas, proporcionando un análisis exhaustivo a partir de bibliografías, de cómo estas tecnologías pueden integrarse de manera efectiva en la gestión documental, destacando sus beneficios, limitaciones y casos de éxito que sirvan para entender cómo se complementa este trinomio, para tener organizaciones eficientes e inteligentes.

Marco de Referencia

Antecedentes

La Ciencia de Datos ha evolucionado significativamente, en la década de los 60 y 70, la cual se centraba en técnicas estadísticas tradicionales aplicadas a conjuntos de datos relativamente pequeños, con la inclusión de la informática y la capacidad de almacenamiento masivo en las décadas siguientes, se hizo posible analizar grandes volúmenes de datos; en la década de 2000, la Ciencia de Datos se consolidó como un campo interdisciplinario que integra estadística, informática y conocimiento de áreas, que abarca desde problemas sencillos, hasta los más complejos.

La Monografía llamada Evolución de la Ciencia de Datos y Big Data en la Gestión Documental, es basada en una recopilación bibliográfica de muchos autores, se resalta el proyecto de maestría (Cuellar, Augusto, Felacio, Agustín J, & Díaz, 2023), que trata de potenciar el proceso de Gestión Documental desde la creación de documentos hasta su disposición final, utilizando el Big Data generado en sus Sistemas de Gestión Documental Electrónica de Archivo (SGDEA), de una forma en que se pueda responder en el menor tiempo posible a los usuarios internos y externos del Servicio Nacional de Aprendizaje (SENA) - Centro Agroecológico y Empresarial en Fusagasugá Cundinamarca.

También el desarrollo de este documento se basa en el boletín de (Díaz, 2022), que expone los retos de integrar la Gestión Documental con las nuevas tecnologías, técnicas de Ciencia de los Datos y el Big Data para recuperar de forma rápida y eficaz la información que se genera en las empresas; la autora habla de “*información inteligente*”.

Se consulta el documento de (Escobar, Mercado, & Rodríguez, 2020), que expone los beneficios de utilizar los datos masivos en las empresas e instituciones gubernamentales

colombianas en todos sus procesos de gestión incluyendo la gestión de información, desde el formato papel, hasta los distintos sistemas de cámaras y sensores para una mejor toma de decisiones; implicando esto en conductas de buen gobierno y lucha contra la corrupción.

Por último, se hace referencia al documento de (De Bustos, 2007), en el que se expone su experiencia en una empresa radiodifusora y de televisión española llamada Telecinco, en un proyecto de digitalización para acercar la tecnología al usuario, dando máxima flexibilidad de acceso a la información en formatos análogos, compartirla y crear un entorno donde la multifuncionalidad sea algo común; donde se comparta la información sin necesidad de moverla para evitar su deterioro. Así las cosas, crearon en la cadena de televisión un repositorio que almacena grandes cantidades de datos indexado a documentos y ficheros multimedia que se actualiza en tiempo real.

Ciencia de Datos

La Ciencia de Datos es una disciplina que combina técnicas de Estadística, Matemáticas e Informática para extraer conocimiento y patrones de los datos, se apoya en una variedad de herramientas y lenguajes de programación, entre los más populares encontramos a Python, R y SQL, es estudiar los datos con el fin de extraer información significativa para generar sabiduría, es plantear y responder a preguntas como ¿Qué pasó?, ¿Por qué pasó?, ¿qué pasará? y qué se puede hacer con los resultados obtenidos (Toro López, 2023)

Big Data

Big Data hace referencia a volúmenes de datos que son demasiado grandes para ser manejados por software tradicional de procesamiento de datos, sus características intrínsecas se conocen como las 5 V's: Volumen, Velocidad, Variedad, Veracidad y Valor; que a partir de estos se puede determinar si las variables para responder preguntas de negocio, se utilizan tecnologías

para procesar los datos como Hadoop y Spark, que son fundamentales en este ámbito para temas de programación, se puede decir que el Big Data o los datos masivos, son el combustible de muchos modelos de aprendizaje automático e Inteligencia Artificial (Marqués, 2015).

Gestión Documental

Gestión Documental es implementar, procedimientos para ordenar, clasificar, describir y conservar la información en diferentes soportes, desde su creación hasta su disposición final; en la era digital, esto incluye la captura, almacenamiento, recuperación y disposición de esos registros, en este proceso, a través de la historia, se plasmaba la información en formato arcilla, pergaminos, papel, cintas, discos, tarjetas hasta tecnologías como los sistemas de almacenamiento en la nube y bases de datos, que son el sostén para alimentar sistemas de aprendizaje automático e inteligencia artificial para la toma de decisiones (Mundet, 2012)

Marco Normativo

Marco Normativo Internacional

- Declaración universal de los archivos - ONU.
- Norma ISO 15489 - Estándar para la gestión de documentos de archivo en cualquier formato o medio
- Norma ISO 30300 - Información y documentación. Sistemas de gestión para documentos. Fundamentos y vocabulario.
- Norma ISO 30301 - Información y documentación. Sistemas de gestión para documentos. Requisitos.
- Norma ISO/IEC 5259 - Modelo de calidad de datos para la analítica de datos basado en aprendizaje automático (Machine Learning - ML).

Marco Normativo en Colombia

- Constitución Política de Colombia. – Artículo 8: Es obligación del Estado y de las personas proteger las riquezas culturales y naturales de la Nación.
- 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.
- Ley 594 de 2000 - Ley General de Archivos.
- Ley 962 de 2005 - Por la cual se dictan disposiciones sobre racionalización de trámites y procedimientos administrativos de los organismos y entidades del Estado y de los particulares que ejercen funciones públicas o prestan servicios públicos.
- Ley 1581 de 2012 - Por la cual se dictan disposiciones generales para la protección de datos personales.
- Ley 1712 de 2014 - Por medio de la cual se crea la Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional y se dictan otras disposiciones.
- Decreto 1080 de 2015 Por medio del cual se expide el Decreto Reglamentario Único del Sector Cultura.
- Decreto 2106 de 2019 - Por el cual se dictan normas para simplificar, suprimir y reformar trámites, procesos y procedimientos innecesarios existentes en la administración pública.
- Acuerdo 001 del 2024 – “Por el cual se establece el Acuerdo Único de la Función Archivística, se definen los criterios técnicos y jurídicos para su implementación en el Estado Colombiano y se fijan otras disposiciones.”

Implicaciones y Usos Prácticos

Mejora de Procesos en Gestión Documental

La integración a través de los años de la Ciencia de Datos, Big Data, Inteligencia Artificial con la Gestión Documental ha demostrado mejorar significativamente la eficiencia en la prestación de servicios archivísticos, elaboración, gestión y disposición final de los documentos, respuesta a incidentes y toma de decisiones. Esto incluye la automatización de procesos repetitivos, la mejora en la recuperación de información y la optimización del almacenamiento de documentos, que actualmente para tomar decisiones en tiempo real los almacenan en grandes bases de datos, extrayendo un petróleo importante que son los metadatos que, es el dato que describe al dato (Huertas Bailén et al,2024).

Toma de Decisiones Basada en Datos

Al hacer ese ejercicio para extraer, transformar, cargar y analizar datos en pequeña o gran escala, permiten a las personas, empresas u organizaciones tomar decisiones más informadas y basadas en datos, sabemos que los documentos sin importar su soporte, están compuestos por datos; esto puede llevar a una mejor gestión del conocimiento e incluso generar sabiduría, también llegar a estrategias más efectivas en la administración de la información, al digitalizar y extraer los registros análogos con sus metadatos y aunarlos con los registros que se producen nativamente de forma electrónica, se podrá tener una visión más amplia de los procesos, problemas de negocio o casos de uso en muchas industrias, desde su creación hasta el presente y de ahí hacia el futuro.

Evolución de la Gestión Documental en la Era Digital: Transformaciones Impulsadas por la Ciencia de Datos y Big Data

La Gestión Documental es un campo que ha experimentado una gran evolución a lo largo de la historia, especialmente con la llegada de la era digital. Desde los años 60 hasta la actualidad, se ha presenciado avances significativos en cómo se almacenan, organizan y gestionan los documentos, antes de la era digital, los documentos se guardaban principalmente en formato papel (Cerrillo Martínez, 2018); desglosándose de la siguiente forma:

Década 1960 – 1970: Fundamentos y Comienzos de la Digitalización

En la década de 1960, consultando el material de AGN - Colombia (2004), surge una tecnología que cambió la forma de como reproducir información, siendo la microfilmación, que se populariza como una de las primeras técnicas para reducir el volumen físico de documentos, estos se reducían a tamaño microscópico almacenándose en rollos de película; ya que para esa época todo se documentaba en formato papel de una forma amplia y necesitaban utilizar un soporte que permitiera albergar un mayor número de documentos en menor espacio.

También en la década de 1970, se implementaron sistemas de información, que, en este caso, las empresas tecnológicas como IBM y Xerox diseñan los primeros sistemas de información computarizada, que permitían almacenar y recuperar información de manera electrónica, pero aún dependían en gran medida de los documentos físicos; así mismo, en el ejército de los Estados Unidos, surge el proyecto llamado ARPANET, que consistía en interconectar las computadoras en una red que permitieran compartir datos entre máquinas de gran tamaño (Cruz Herradón, 2013).

Tukey (1962), publicó un artículo titulado "*The Future of Data Analysis*", marcando un punto de inflexión en la historia de la estadística y la ciencia de datos, se redefinió cómo se

percibía esta disciplina, argumentando que el análisis de datos debería ser considerado como un campo independiente, separado de las matemáticas puras y de la estadística teórica; para destacar la importancia de centrarse en la práctica del análisis de datos para resolver problemas que se presenten en varios aspectos de la vida cotidiana.

Tukey (1977), publica su libro, donde explica la importancia de introducir un enfoque revolucionario para analizar datos, el destaca la importancia de explorar los datos de manera visual e intuitiva antes de aplicar métodos estadísticos formales; además propuso herramientas como gráficos de caja (boxplots), diagramas de dispersión y otras técnicas gráficas para identificar patrones, tendencias y anomalías. Con la finalidad de que fuera aplicado transversalmente a cualquier tipo de disciplina y/o ciencia.

En temas de gestión documental, para esta década, muchas empresas e instituciones, optaban por utilizar sistemas de gestión documental análogos, pero en la primera mitad del siglo XX se caracterizó por el nacimiento de la historia científica con la *École des Annales*, la cual impulsó el uso de los archivos por parte de los historiadores. Los nuevos procesos de producción y reproducción de documentos, el ritmo de las sociedades industriales, entre otros aspectos, planteó un problema: la cantidad de masas documentales cuya conservación resultaba imposible y sobre todo, un gran dilema, esa fuente de información era necesaria para las administraciones, las cuales no necesitaban historiadores, sino técnicos capaces de gestionar toda esa documentación (Baratz, 2021).

Para inferir: La evolución tecnológica desde la década de 1960 ha transformado radicalmente la manera en que se maneja, almacena y analiza la información sin importar su soporte, la microfilmación marcó un hito al reducir el volumen físico de los documentos, mientras que los sistemas de información computarizada en los años 70, como los desarrollados

por IBM y Xerox, comenzaron a integrar métodos electrónicos para la gestión documental, aunque aún dependían del papel, paralelamente, surge la ARPANET, que sentaron las bases de la interconectividad digital que se conoce hoy en día.

En el ámbito del análisis de datos, Tukey revolucionó la disciplina al proponer un enfoque más práctico e intuitivo, destacando la exploración visual de los datos y estableciendo su independencia de la estadística teórica, estos conceptos han trascendido hacia aplicaciones en diversas disciplinas y casos prácticos.

En cuanto a la gestión documental, las necesidades crecientes de las sociedades industriales impulsaron nuevas formas de organización y preservación de archivos, buscando no solo conservar grandes volúmenes de información, sino también optimizar su uso para fines administrativos, legales, técnicos e históricos; donde fue necesario preguntarse: ¿Quiénes debían ser los idóneos para administrar información? Bajo este panorama, se evidencia cómo los avances en tecnología y ciencia de datos han sido fundamentales para abordar los retos asociados al manejo de grandes cantidades de información en diversos contextos y que haya personas capacitadas para permitir el acceso a la información, que está compuesta por datos.

Década de 1980: Era de las “Personal Computers” e Institucionalización de los Archivos

El Archivo adquiere la categoría de servicio dentro de las instituciones y de una postura estática e inmovilista tiende a la dinamización buscando una proyección cultural con un firme propósito de integración. El reconocimiento de servicio muchas veces más teórico que real, lo llevara a formar parte de los proyectos culturales y científicos de la Política de cualquier país e incluso de las divisiones administrativas en que estén organizados como es el caso de nuestras Autonomías (Heredia, 1986).

La computación experimentó una revolución, empezando con la introducción de las computadoras personales, como el Apple II, y la IBM PC, que transformaron radicalmente la forma en que las personas interactuaban con la tecnología, llevando esto del ámbito empresarial y académico a los hogares; multinacionales como Microsoft y Apple se convirtieron en gigantes tecnológicos, con el lanzamiento de sistemas operativos como MS-DOS y el revolucionario Macintosh con su interfaz gráfica de usuario; el desarrollo de microprocesadores más potentes, como el Intel 8086 y 80286, que permitieron computadoras más rápidas y compactas. Además, surgieron lenguajes de programación como BASIC, y se sentaron las bases para las futuras tecnologías de comunicación e información, que transformaron la forma de cómo se gestionaba la información en soportes físicos y digitales (Muñoz M., 2009).

Con la utilización de las computadoras personales, en entornos empresariales y personales, se necesitaba procesar la información y posteriormente almacenarla para posteriores consultas, pero en vista de que no estaba tan desarrollada la internet ni los sistemas avanzados que se conocen actualmente estos datos e información, eran impresos en su época en formato papel, lo que implicaba un gran problema como producción descontrolada, aumento de versiones, multiplicación de copias, gavetas o almacenes desordenados y difícil acceso, aumentando la pérdida de información. A reglón seguido, se suma que las empresas u organizaciones no valoraban ni conocían la importancia de tener un profesional gestor de grandes cantidades de información y ni conocían tampoco el concepto del Big Data (Russo, 2013).

Para la época, no se hablaba en sí de “Gestión Documental” simplemente denominaban estas técnicas, procesos y procedimientos administrativos para gestionar la información como archivística y los lugares donde los almacenaban, se llamaban “Archivos”. También se acuñó el concepto de archivo, como un contenedor de información, que era almacenado en medios

computarizados, que son los datos representados en grafos, letras, palabras o símbolos que los componen (Sainz, 2023),

Las empresas querían hacer cálculos complejos, por lo cual se tomaba mucho tiempo hacer este tipo de procedimientos a mano y almacenarlos en cuartos con una amalgama de información; además hacer las gráficas correspondientes, para que los directivos de las empresas observaran en un solo vistazo las métricas requeridas para la toma de decisiones. Fue cuando en 1985 lanza la empresa Microsoft su hoja de cálculo Excel, que permite a los usuarios crear hojas de cálculo o presentaciones de filas y columnas en las que se registran datos. Sumando a ello, se podía hacer aritmética y cálculos complejos, coadyuvando en la forma de como se almacenaba la información, ya que esta se guardaba en la computadora y no en los rollos tradicionales de papel (Mark, 2024).

Se infiere, que la evolución de la gestión de información ha estado profundamente influenciada por los avances tecnológicos y su impacto en las dinámicas empresariales y sociales, Desde la concepción inicial de los archivos como contenedores de información a instituciones que presten un servicio social, la introducción de las computadoras personales, transformó la interacción con la tecnología y sentó las bases para la digitalización y el surgimiento de herramientas como Excel, que redefinieron el almacenamiento y procesamiento de datos. Sin embargo, el contexto inicial presentaba limitaciones significativas, como la dependencia del papel y la falta de valorización de la información estratégica, lo que generaba desorden y pérdida de datos. Este desarrollo histórico revela cómo la gestión documental, aunque no se conceptualizaba como tal, evolucionó, marcando un antes y un después en la forma en que se administra y aprovecha la información en diversos entornos.

Década 1990 – 2000: Internet y Globalización

En la década de 1990, expone Chávez (2018), que la ciencia de datos, internet y la globalización experimentaron una transformación radical con el nacimiento de la World Wide Web, democratizando el acceso a la información y conectando globalmente a millones de personas; estas empresas comenzaron a comprender el valor estratégico de los datos, pero no sabían cómo gestionarlos.

Los expertos en ciencias de la computación, acuñaron en 1999 el concepto de Big Data, para poder administrar grandes y complejos conjuntos de datos, que no podían procesarse con métodos tradicionales, los cuales siempre han contenido características muy diferentes en tipo, velocidad, variedad e incluso veracidad; en lo cual con el transcurrir del tiempo, se desarrollaron técnicas iniciales de minería de datos o algoritmos para ranquear y recuperar la información que se iba organizando, exponiendo el caso de Google (fundado en 1998) empezaron a visionar como organizar la información mundial que estaba en libros, documentos, discos de audio, video casete entre otros (Castillo, 2019).

Hace muchos años, el presidente estadounidense Truman creó la Comisión Hoover, con el propósito de encontrar una solución para la conservación de grandes volúmenes documentales, surgidos como consecuencia del reordenamiento de la administración norteamericana durante la Segunda Guerra Mundial; de esta forma se producía, se reconocía y se aceptaba una demarcación rigurosa entre la labor de los archiveros y los profesionales que se dedicarían a la gestión de los documentos, los primeros trabajarían con la documentación denominada activa o semiactiva y los segundos con aquella que fuera a ser conservada permanentemente. La demarcación no comenzaría a ser superada hasta entrada la década de los años noventa del siglo XX, e incluso en muchos países aún se mantiene; sin embargo, para que el trabajo fuera organizado, se crearon los

tres tipos de archivos: Archivo de Gestión, Archivo Central, Archivo Histórico (Mena Mujica, 2005).

En el año 2006, se crea una potente herramienta para administrar el Big Data llamada Hadoop, que según White (2015), su funcionalidad es para programar aplicaciones distribuidas que manejen grandes volúmenes de datos (big data) permitiendo trabajar con miles de nodos (computadoras) en red y Petabytes de datos, inspirándose en los documentos de Google sobre MapReduce y Google File System (GFS).

Por otro lado, en el año 2008, utilizando el lenguaje de programación Python, para procesar datos a pequeña escala, se desarrolló una librería que cambiaría el rumbo de la historia, la cual se llama Pandas, acrónimo de Python Data Analysis; además en el mundo de la Ciencia de los Datos, empezaron analizar la importancia de mezclar la estadística con la computación, para abarcar problemas transversalmente a cualquier proceso empresarial (Maldonado, 2022).

En conclusión, la evolución de la ciencia de datos y el Big Data ha sido un proceso continuo y transformador, impulsado por avances tecnológicos, la globalización y la necesidad de gestionar grandes volúmenes de información; Desde los primeros esfuerzos por organizar datos en la década de 1990 con la aparición de la World Wide Web y la conceptualización del Big Data en 1999, hasta el desarrollo de herramientas como Hadoop en 2006 y Pandas en 2008, se ha demostrado cómo la integración de técnicas estadísticas, computacionales y documentales ha permitido enfrentar retos crecientes en la gestión de datos e información en cualquier tipo de soporte, suscitada por un proceso. En la Gestión Documental, estos avances han marcado un punto de inflexión en la manera de cómo las empresas u organizaciones valoran y procesan este activo, destacando su impacto estratégico en un mundo cada vez más interconectado para facilitar la consulta o conservación a largo plazo.

Década de 2010: Computación en la Nube, Digitalización y Ética en la Gestión Empresarial

En la década del 2010, la Ciencia de Datos, la computación en la nube y la virtualización revolucionaron radicalmente el proceso de gestión documental, transformando por completo cómo las organizaciones almacenan, procesan, acceden y analizan la información (Celaya Luna, 2014).

Estas tecnologías permitieron migrar de sistemas tradicionales de archivos físicos a plataformas digitales inteligentes, donde el big data y el Machine Learning posibiliten a futuro la clasificación automática de documentos, la extracción de información relevante mediante algoritmos avanzados y el almacenamiento escalable y seguro en infraestructuras virtuales. La nube facilitó el acceso remoto y colaborativo, reduciendo costos de infraestructura física y mejorando la eficiencia, mientras que las técnicas de ciencia de datos introdujeron capacidades de búsqueda predictiva, análisis de contenido y gestión inteligente de documentos, convirtiéndolo que antes era un proceso administrativo complejo, a un sistema dinámico, interconectado y altamente optimizado que pueda adaptarse rápidamente a las necesidades cambiantes de las empresas modernas (Carrillo, 2008).

Para esta década, en temas de Gestión Documental, se manifestó la necesidad de implementar en muchas empresas e instituciones públicas, la organización de toda la documentación producida a través del tiempo, aplicando instrumentos archivísticos como Tablas de Retención Documental , para seleccionar la información producida que se va a transferir de los Archivos de Gestión al Central y del Central al Histórico con sus respectivas copias de seguridad en soportes digitales; todo ese desorden documental, ya que producto de la indiferencia, corrupción y otros factores que generaron que las empresas perdieran valor en la elaboración de sus productos o prestación de servicios (Vivas et al., 2013).

Para conservar la documentación a largo plazo, se implementaron en muchas empresas e instituciones, el cargue de todos los documentos físicos de la empresa a servidores o centros de datos, para recuperar los documentos previamente organizados, con la finalidad de recuperar en tiempo récord por parte de los administradores de información y responder a diferentes requerimientos por parte de personas naturales o jurídicas; también aparece el concepto de documento electrónico y los Sistemas de Gestión Documental electrónica de archivo (SGDEA) (García Morales, 2013).

Emprendieron muchas industrias y negocios, digitalizar o migrar los Formatos Únicos de Inventario Documental (FUID) a sistemas capaces de mostrar datos relevantes como número de folios, tomos, carpetas, cajas, legajos, tipo de soporte, fechas extremas entre otros (Ibañez Montoya, 2008).

En conclusión, durante la década del 2010, la Ciencia de Datos, la computación en la nube y la virtualización transformaron radicalmente la gestión documental, permitiendo a las organizaciones migrar de sistemas físicos tradicionales a plataformas digitales más eficientes e inteligentes; estas tecnologías facilitaron el almacenamiento escalable, el acceso remoto, la clasificación automatizada y el análisis avanzado de documentos mediante algoritmos de Big Data y Machine Learning. Sin embargo, la necesidad de aplicar instrumentos archivísticos, como las Tablas de Retención Documental, y organizar el desorden documental persistente en muchas instituciones y empresas, fue necesario permear procesos para asegurar la preservación, recuperación y uso eficiente de la información.

La digitalización de archivos y el uso Formatos Únicos de Inventario Documental, han optimizado la gestión, contribuyendo a la mejora de procesos y al aumento del valor de los

servicios y productos ofrecidos por las organizaciones en materia administración de la información.

Década de 2020: Inteligencia Artificial y los Datos como su Combustible

En la actualidad, la ejecución de la Inteligencia Artificial (IA) como una tecnología central en diversos sectores económicos y sociales, actualmente, los datos han adquirido el estatus de el “nuevo petróleo”, siendo este un insumo esencial para el desarrollo y entrenamiento de estos modelos; el avance en herramientas de análisis, aprendizaje automático así como redes neuronales, ha sido exponencial, ya que permite transformar las aplicaciones desde la atención médica personalizada hasta la automatización industrial, incluyendo los procesos de organización, consulta y conservación documental. Tecnologías como el aprendizaje profundo, combinadas con grandes volúmenes de datos provenientes de dispositivos IoT (Internet de las cosas), redes sociales, sistemas empresariales y documentos digitalizados, que han hecho posible predecir comportamientos, optimizar recursos y resolver problemas complejos a gran velocidad (Bazán Gil, 2023).

Para cumplir con ese interrogante escrito en la justificación de la monografía, consistente en aprovechar estas tecnologías que ayudan almacenar, procesar, analizar y tomar decisiones a la hora de abarcar un desafío archivístico, que beneficia procesos tediosos como consultar en el menor tiempo posible un documento o libro, para responderse a un peticionario, también se utiliza el Big Data y la Inteligencia Artificial, para realizar procedimientos de disposición final en documentos, cuantas veces se consulta un fondo documental, así como en qué condiciones de temperatura y humedad relativa se conserva estos activos (Díaz, 2022).

La capacidad de almacenar, procesar y analizar datos masivos, aprovechando la infraestructura de computación en la nube, sigue siendo un pilar fundamental en los modelos de

inteligencia artificial, no solo consume datos para generar conocimiento y valor, sino que también produce nuevas fuentes de información, alimentando un ciclo de innovación constante, para mejorar procesos de la Gestión Documental, desde la planeación hasta la consulta, conservación y/o disposición final por parte de personas naturales o jurídicas (GESDA, 2017).

La sinergia, entre datos que están plasmados en diferentes soportes y los algoritmos en la actualidad, han redefinido procesos e industrias y plantean otra circunstancia referente a los desafíos éticos y técnicos sobre la privacidad, la seguridad y la gobernanza de la información, marcando un antes y un después en la evolución tecnológica global (Mojapelo y Marutha, 2023).

A lo expuesto en los párrafos anteriores, se añade una paradoja clave en torno a la privacidad de los datos masivos (Big Data) que alimentan los sistemas de inteligencia artificial y las personas que utilizan las herramientas de Ciencia de Datos, para gestionar el activo y generarlo como conocimiento y es el ejercicio del derecho al acceso de la información y la protección de esos datos en sí, ya que se deben crear protocolos para el acceso a la información en caso de que tenga alguna confidencialidad, para que personas inescrupulosas tengan acceso restringido y no puedan ser difundidos (Villar, 2021).

Se infiere que la Inteligencia Artificial (IA) y el Big Data son tecnologías esenciales que transforman muchos sectores incluyendo la gestión documental, al aprovechar grandes volúmenes de datos para optimizar recursos, automatizar procesos y resolver problemas complejos; los datos, considerados el "nuevo petróleo", alimentan modelos avanzados que generan conocimiento y valor, aunque plantean desafíos éticos y técnicos relacionados con la privacidad, la seguridad y la gobernanza de la información. En la gestión documental, estas tecnologías mejoran desde la consulta hasta la conservación de documentos, exigiendo

protocolos para equilibrar el acceso a la información con su protección frente a usos indebidos, en un contexto de constante innovación y evolución tecnológica.

Herramientas y Técnicas de Ciencia de Datos y Big Data en la Gestión Documental: Impacto en la Eficiencia y Optimización de Procesos

La gestión documental, como proceso fundamental en las organizaciones, ha evolucionado significativamente con la incorporación de herramientas y técnicas avanzadas de Ciencia de Datos y Big Data. En un entorno donde los volúmenes de información crecen exponencialmente, estas tecnologías permiten optimizar cada etapa del ciclo documental, desde la elaboración, captura y almacenamiento hasta la clasificación, recuperación, conservación y eliminación de documentos (Ríos y Fraile, 2020).

La integración de métodos como el procesamiento de lenguaje natural (NLP), algoritmos de aprendizaje automático y bases de datos no relacionales ha transformado la manera en que se manejan los datos documentales, mejorando la eficiencia operativa y administrativa, reduciendo costos; en este contexto, el análisis y aprovechamiento estratégico de grandes volúmenes de datos se convierten en una ventaja competitiva clave para las organizaciones, permitiéndoles tomar decisiones basadas en información estructurada y precisa (Viloria, 2023).

Figura 1

Organización y Digitalización de Archivos



En la anterior ilustración, se puede evidenciar que los formatos en los Archivos, los cuales han sido tecnificados por procesos de gestión documental, donde la información física, se ha migrado a repositorios digitales; en los cuales se almacenan metadatos que se puede utilizar mediante procesos de ciencia de datos, para la mejora en la toma de decisiones como la cantidad de documentos que caben en una estantería, para optimizar el almacenamiento en las bodegas de Archivo (Martínez, 2023).

Esto coadyuva en que las consultas se hagan en el menor tiempo posible, para responder los diferentes requerimientos por parte de personas naturales y jurídicas, garantizando ese derecho humano de acceder al conocimiento (Téllez, 2020).

En la siguiente tabla, se puede evidenciar que, en el proceso de Gestión Documental, se utilizan herramientas y el impacto que esta genera; basado en la información consultada en el material de (Xu, 2022):

Tabla 1*Uso de Herramientas y Tecnologías para el Proceso de Gestión Documental*

Proceso de Gestión Documental	Año	Herramientas	Descripción	Impacto	Norma ISO
Captura y Digitalización	1980	Software OCR	Convierte documentos físicos en digitales y extrae datos clave automáticamente	Digitalización rápida y precisa; elimina la dependencia de documentos físicos	ISO 30301
Consulta	2002	Gestor documental ORFEO	Facilita la gestión de documentos en cualquier empresa.	Centralizar y organizar toda la información en una plataforma digital	ISO 30300
Almacenamiento y organización	1970	Bases de Datos SQL	Almacena grandes volúmenes de datos no estructurados en repositorios accesibles y organizados	Gestión centralizada y accesible; optimización en la recuperación de información, para hacer predicciones.	ISO 15489
	1990	Bases de datos No SQL			
	2006	Hadoop			
	2013	Databricks			
Clasificación y Procesamiento	1960	Lenguaje de procesamiento Natural – NLP	Analiza texto para categorizar documentos y predecir patrones en su uso	Automatización de la clasificación; detección de tendencias en el flujo documental	ISO/IEC 5259
	2018	Algoritmos de Machine Learning			
Control / Seguimiento	2015	Power BI	Monitorea el flujo documental, genera reportes dinámicos y automatiza procesos de control	Supervisión eficiente; reducción de errores y alertas oportunas para auditorías	ISO 30301
	2003	Tableau			

La tabla presentada anteriormente, detalla las diversas herramientas y tecnologías aplicadas en el proceso de Gestión Documental, abarcando desde la captura y digitalización hasta el control y seguimiento; cada herramienta, asociada a un año de implementación, resalta su impacto en la mejora de la eficiencia y accesibilidad del flujo documental, así como su contribución al cumplimiento de normas ISO relevantes. Este análisis evidencia cómo la evolución tecnológica, como el uso de bases de datos, NLP y visualización de datos, ha transformado y optimizado la gestión documental, alineándola con estándares internacionales para garantizar su eficacia y adaptabilidad en entornos organizacionales.

Beneficios, Desafíos y Casos de Estudio en la Implementación de Ciencia de Datos y Big Data en la Gestión Documental

En el contexto de la implementación de Ciencia de Datos y Big Data en la gestión documental, este apartado se centra en analizar los beneficios, desafíos y casos de estudio significativos en tres organizaciones: LAIEX, Policía Nacional de Colombia y Google; Cada una de estas entidades representa enfoques únicos hacia la transformación digital y la optimización de sus procesos documentales. LAIEX ha demostrado cómo la digitalización y el análisis de datos pueden optimizar procesos operativos en el sector industrial; la Policía Nacional de Colombia ha implementado sistemas avanzados para fortalecer la accesibilidad de la información; mientras que Google lidera con herramientas de inteligencia artificial aplicadas a la gestión documental, estableciendo nuevos estándares de innovación global.

Empresa LAIEX

LAIEX es una empresa del sector químico que entró en funcionamiento desde el año 1970, dedicada a la fabricación de compuestos derivados del caucho, para la industria de las moquetas en los sectores de automóvil, construcción y textil, desde el año 2000 amplió su portafolio de productos en el campo de la polimerización de emulsiones acrílicas y vinílicas en base acuosa para el sector de los recubrimientos de pintura, calzado y construcción. Contando con más de 200 soluciones, entre productos, servicios y tecnologías (Marta, 2020).

La principal dificultad a la que se enfrentaba era la gestión de su documentación, que llevaba mucho tiempo y obligaba a imprimir muchos documentos, especialmente en los procesos de facturación (compra y venta), lo cual suscitaba un agravamiento al tratarse de una empresa que estaba en expansión, lo que triplicaba copias de ejemplares, ya que aumentaban cada vez los proveedores o clientes, en este caso, Europa y América. En estos casos, hace que la empresa

entre en una situación de caos e ineficiencia en sus funciones operativas como administrativas; ya que no había una planificación estratégica en el ciclo de la producción documental, sea física, digital o electrónica (Borras, 2021).

Otra situación para sustentar, eran las condiciones que dificultaba la gestión de los documentos, era su dispersión, al trabajar en tres sedes ubicadas en tres países diferentes y separadas por miles de kilómetros salía costoso el envío de copias, afectando las finanzas de LAIEX, por lo cual, contrataron a la consultora española GDX Group, encargada de emplear soluciones de transformación digital, para gestionar la información de forma eficiente (Dominguez, 2021).

Así las cosas, detectaron que se requería por, ejemplo, tener mejor control de las facturas en el proceso de compras, puesto que la validación de los documentos dependía de varios responsables que no se encontraban físicamente en el lugar de los mismos, en cuanto al proceso de ventas, también era necesario automatizar toda esa gestión, ya que esta se realizaba de manera manual y se requería la impresión de muchos documentos, generando alto consumo de papel y tinta; además había que hacer llegar esta información actualizada a los clientes o proveedores, desplegando también servicios de logística o envío de paquetería. Analizándolo desde una perspectiva más amplia esto afectaría el valor de una empresa u organización, ya que la información no se consolida de forma centralizada y tendrán que sacar copias para enviar a diferentes partes para su análisis y validación en muchos casos (García , 2016).

Para solucionar este problema, LAIEX empleó un sistema de digitalización de documentos físicos, también dentro del misma herramienta, un módulo de facturación electrónica, donde según el rol cada colaborador, se puede acceder desde cualquier lugar, para visualizar y aprobar dichos documentos y/o facturas; estos metadatos, se extraen, procesan y se

almacenan en un DataWare- House para la tomar decisiones en los procesos de ventas, compras y mejoras en procesos de elaboración de sus productos, esto coadyuva a mejorar las ventas, gestionar clientes y desarrollar de nuevos modelos de negocio. Al utilizar procesos de transformación digital, gestión del Big Data y utilizar las técnicas en Ciencia de Datos, mejora la eficiencia de la empresa, los costos y gastos, así mismo, ver las necesidades o tendencias que tienen los stakeholders, pero también ayuda a una buena gestión de la información evitando perdida, duplicidad y autenticidad de la información (Sanabria et al, 2022).

Para Inferir, la experiencia de LAIEX demuestra cómo la transformación digital, desplegada en aplicativos, Ciencia de Datos y Big Data, no solo resuelve problemas operativos, sino que se convierte en un catalizador de cambio estratégico para las organizaciones, la digitalización de procesos críticos como la gestión documental y la facturación no solo optimizó recursos y redujo costos, sino que también permitió una integración más eficiente de la información sin importar su soporte, facilitando la toma de decisiones informadas y el desarrollo de nuevos modelos de negocio. Este caso subraya la importancia de adaptarse a las exigencias del entorno globalizado, donde la innovación tecnológica es clave para mantener la competitividad, mejorar la relación con las partes interesadas y garantizar la sostenibilidad a largo plazo en un mercado en constante evolución.

Policía Nacional de Colombia

Consultado la web de (Policía Nacional, s.f.), es una institución centenaria, que fue creada por medio del decreto 1000 del 5 de noviembre de 1891, cuando el señor Carlos Holguín era el presidente de Colombia, organiza un cuerpo de Policía Nacional, el nuevo cuerpo de Policía pasó a depender del Ministerio de Gobierno sin perjuicio de recibir instrucciones del Gobernador de Cundinamarca y alcalde de Bogotá. Luego de muchos años, en la guerra

bipartidista, los agentes de la época tenían filiaciones políticas hacia los bandos liberal o conservador, cuando ocurre el famoso “Bogotazo” esta institución se dividió en dos bandos al servicio de los políticos de ese entonces; luego de ese hito, por órdenes del señor General Gustavo Rojas Pinilla pasa a ser orgánica del ministerio de guerra quedando como una Policía militarizada, en la cual documentaban sus actividades en libros y pergaminos.

Al convocarse la constituyente, por la cual se sanciona la constitución política de 1991, la cual define un “estado social de derecho”, la policía nacional según (Senado Colombia, 1991), en su artículo 218 señala que: “La Policía Nacional es un cuerpo armado permanente de naturaleza civil, a cargo de la Nación, cuyo fin primordial es el mantenimiento de las condiciones necesarias para el ejercicio de los derechos y libertades públicas, y para asegurar que los habitantes de Colombia convivan en paz.” Así las cosas, al cambiar sus funciones, cambia su funcionamiento operativo y administrativo, suscitando un despliegue de macroprocesos, que se soportaban en diferentes formatos, como papel, discos flexibles (Disquete), discos duros, cinta de microfilm, acetatos y libros, adicionando todos los documentos históricos en pergaminos, los cuales no se encontraban en condiciones adecuadas para su conservación y recuperación.

Esta situación después de muchos años se suscitó un caos administrativo, ya que, al no tener repositorios de información debidamente organizados, sin los metadatos adecuados y no tener personal idóneo en este tipo de disciplinas para gestionar adecuadamente esos activos, en muchos casos se perdían desde demandas de reparación directa hasta la pérdida de la confianza por parte de los ciudadanos (Leal Buitrago, 2015).

En el año 2010, el director de la Policía en esa época, propuso una transformación de la institución, la cual fue aceptada por el presidente de la república en ese entonces, para incluir y/o implementar nuevas tecnologías computacionales, correo electrónico institucional, además del

Sistema de Gestión Documental, del cual se desplegó instrumentos archivísticos como Tablas de Retención Documental (TRD), Tablas de Valoración Documental (TVD) e Inventarios documentales (FUID), así como la implementación de formatos estandarizados para tener descrita y asequible la información en todas las fases del Archivo de todas unidades de policía a lo largo y ancho del país (Carvajal, 2011).

Para el año 2021, se propone por directivas presidenciales, una transformación policial en la institución, en la cual incluían la profesionalización del servicio de policía basada en los Derechos Humanos, la gestión adecuada de los recursos, de dicha transformación, arrojan resultados como la implementación de un Sistema de Gestión Documental Electrónica de Archivo, llamado Gestor de Documentos Policiales GEPOL, el cual utiliza una gran base de Datos relacional, donde se registran los documentos que son producidos y los que se reciben en la institución; consultado el documento de (AGN - Colombia, 2016), en su tiempo de tantos sistemas heredados, los cuales fueron fallidos para gestionar grandes cantidades de información con sus respectivos metadatos, surge el desarrollo de GEPOL, el cual fue implementado totalmente en la institución, mediante la resolución No. 01000 de fecha 31 de marzo del 2021.

Figura 2

Acceso al Gestor de Documentos Policiales - GEPOL



Nota. Tomada de. Policía Nacional de Colombia (2021)

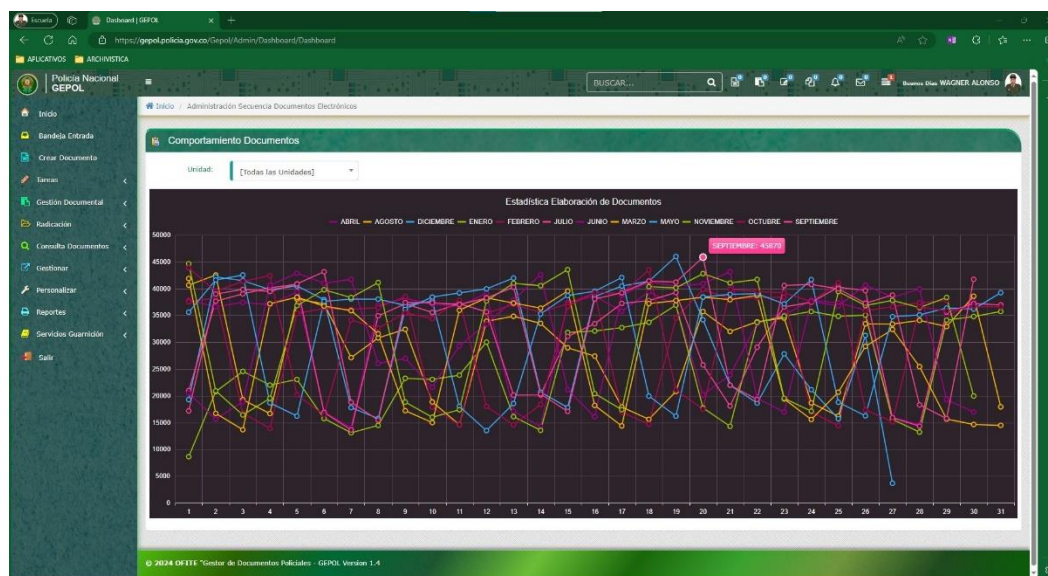
También se consolidó en la mayoría de las unidades policiales de todo el país, el Sistema de Gestión Documental, adicionando como esteroide el SGDEA – GEPOL, del cual se puede consultar en un tablero de visualización, que muestra la cantidad de documentos producidos y recibidos, a nivel general o por unidades policiales.

Consultando el documento (Policía Nacional, 2023), se implementa un Plan Estratégico Institucional - PEI, donde en uno de sus objetivos estratégicos está escrito que se debe “Promover la transformación digital y el fomento de la cultura de cambio e innovación, fortalecimiento ciberseguridad”; en lo cual se empezó crear lagos de datos, para unificar las diferentes Bases de Datos que integran los sistemas de información que coadyuvan en la parte Operativa, Administrativa y Docente de la institución, incluyendo GEPOL, los cuales se conectan a Tableros de visualización, Sistemas Inteligencia Artificial o Aprendizaje Automático, para anticipar hechos que afecten la convivencia y seguridad ciudadana, incidentes que puedan afectar la ciberseguridad o hitos que puedan generar daño antijurídico. Además, gestionar de forma eficiente el servicio monopolizado de Policía.

Aunando lo anterior al utilizar ese gran lago de datos, se han abarcado temáticas como la producción eficiente de documentos, reducción en el consumo del papel, con base a datos estadísticos en los diferentes sistemas, se ejecutan procesos de búsqueda y recuperación de información de una forma rápida, para atender los requerimientos por parte de los usuarios internos y externos, que ha generado un aumento en la credibilidad y confianza de los ciudadanos, a pesar de que hay cosas por mejorar. Sin embargo, se proyecta al año 2050 una Policía más preparada para enfrentar el cibercrimen, con ayuda de Inteligencia Artificial, la cual detrás de ella están los datos y documentos que con estos avances pioneros en su campo, se pronosticará una efectividad en sus funciones (Mendez, 2024).

Figura 3

Tablero Comportamiento Documentos



Nota. Tomada de. Policía Nacional de Colombia (2021)

Se concluye que la Policía Nacional de Colombia, ha experimentado una transformación significativa desde su creación en 1891, inicialmente, la institución enfrentó desafíos administrativos y operativos, especialmente durante la guerra bipartidista y el "Bogotazo"; La falta de un sistema de Gestión Documental adecuado, resultó en la pérdida de información crucial y la disminución de la confianza ciudadana. Sin embargo, con la implementación de Nuevas Tecnologías y Sistemas de Gestión Documental desde 2010, y la posterior profesionalización del servicio aplicada desde 2021, la Policía Nacional ha logrado mejorar significativamente su eficiencia y transparencia.

La adopción del Sistema de Gestión Documental Electrónica de Archivo (SGDEA), desplegada en el Gestor de Documentos Policiales (GEPOL), ha permitido una mejor organización y acceso a la información, lo que ha fortalecido la ciberseguridad y la capacidad de

respuesta de la institución ante los desafíos del país, la creación de lagos de datos y el uso de inteligencia artificial han optimizado la producción y recuperación de documentos, reduciendo el consumo de papel. A pesar de los avances, aún existen situaciones que se deben mejorar, pero los esfuerzos realizados han sentado una base sólida para una gestión más eficiente y confiable en el futuro.

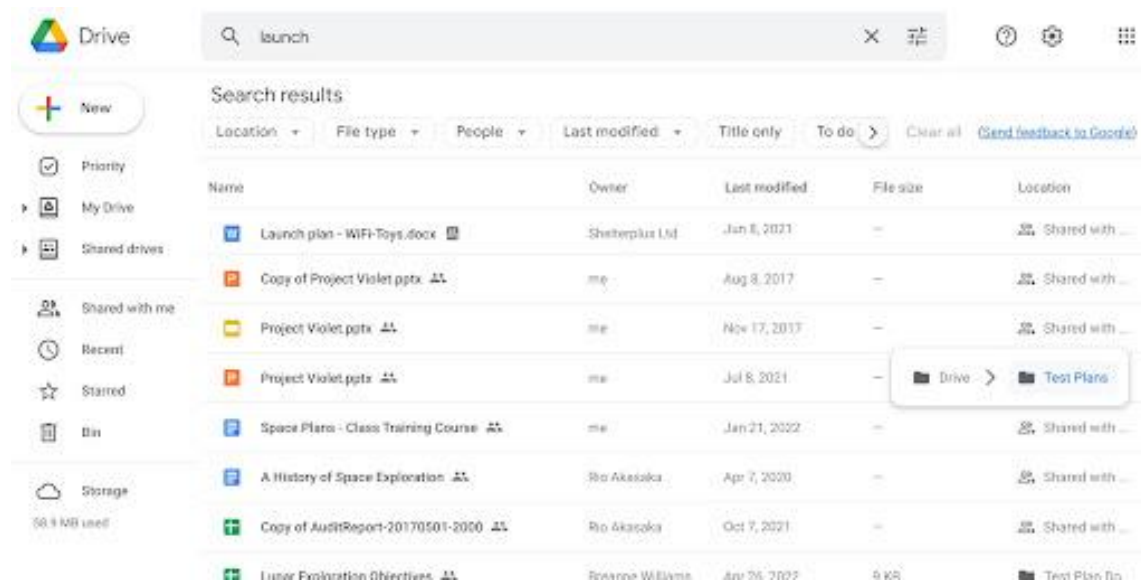
Google

Desde su fundación en 1998 por Larry Page y Sergey Brin, según (Miroslav, 2024) Google se ha destacado por su enfoque innovador hacia la tecnología, aunque inicialmente su objetivo inicialmente era organizar la información en la web a través de su motor de búsqueda, su crecimiento exponencial les planteó el desafío de gestionar de manera eficiente sus propios datos y documentos internos. En sus primeras etapas, Google operaba con un volumen creciente de documentos relacionados con contrataciones, investigaciones y estrategias empresariales, como muchas empresas nacientes, comenzó utilizando métodos tradicionales de archivo en papel; sin embargo, a medida que la compañía iba creciendo y expandiendo sus operaciones a nivel global, se hizo evidente que estos métodos eran ineficientes, costosos y poco escalables.

La transformación comenzó con la digitalización masiva de documentos físicos, un proceso que no solo implicó escanear archivos, sino también implementar estrategias para estructurar, indexar y recuperar documentos rápidamente, esto marcó el inicio del uso de herramientas digitales básicas, pero la verdadera innovación llegó con la integración de Ciencia de Datos y Big Data en sus procesos de gestión documental. En 2012, Google lanzó Google Drive, una plataforma que revolucionó la gestión documental al proporcionar almacenamiento en la nube y colaboración en tiempo real, permitiendo almacenar y compartir documentos de manera eficiente (Rodríguez de Sepulveda Mailillo, 2016).

Figura 4

Captura de Pantalla Aplicación Google Drive



Nota. Tomada de. Arroba System (2022)

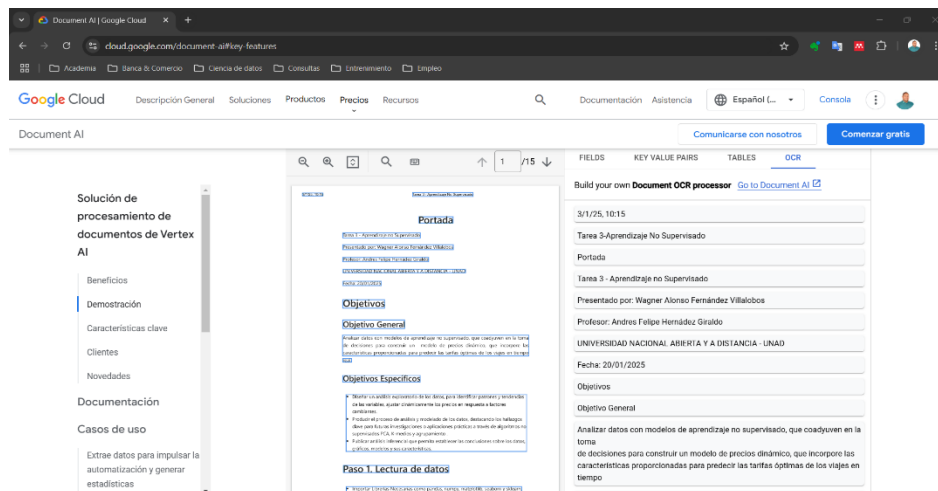
Google Drive, sirvió como la base de una estrategia de gestión documental que unió almacenamiento masivo con accesibilidad universal. permitiendo a los equipos internos trabajar en proyectos globales sin las barreras impuestas por los sistemas tradicionales de almacenamiento en servidores locales. En su compromiso por mejorar continuamente, Google desarrolló en el año 2020 Document AI, una solución basada en Big Data y aprendizaje automático diseñada para automatizar la extracción de datos estructurados y no estructurados. Esta herramienta resultó especialmente útil para procesar grandes volúmenes de contratos, facturas y otros documentos corporativos (Gonzalo Edesa, 2018).

Document AI permite no solo leer texto, sino también entender el contexto y categorizar información clave, mejorando procesos críticos como auditorías internas, manejo de riesgos y toma de decisiones basada en datos. Además, con la capacidad de manejar múltiples idiomas,

Document AI se convirtió en un recurso esencial para las operaciones globales de Google (Hegghammer, 2021).

Figura 5

Uso Herramienta Document AI



Nota. Tomada de. Google (s.f.)

Se concluye que la evolución de la gestión documental en Google es un ejemplo claro de cómo la Ciencia de Datos y el Big Data pueden transformar procesos críticos en una empresa, herramientas como Google Drive y Document AI no solo han mejorado la eficiencia y la accesibilidad, sino que también han establecido un estándar de innovación en la industria.

Con su enfoque en la automatización y el análisis inteligente de documentos, Google continúa liderando el camino hacia un futuro donde la gestión documental no es solo una tarea administrativa, sino una fuente estratégica de valor empresarial; que ha sido un referente para ofrecer servicios y soporte a otras empresas e industrias, con la finalidad de conservar activos de información para la toma de decisiones, transmitir conocimiento y mantener viva la memoria histórica.

Conclusiones

La Ciencia de Datos, el Big Data y el proceso de Gestión Documental se interrelacionan para transformar tanto el ámbito personal como el empresarial, marcando un antes y un después en la forma en que se gestionan y preservan documentos y datos. La incorporación de técnicas avanzadas de análisis de datos, especialmente a partir de metadatos, y el aprovechamiento de grandes volúmenes de información han permitido optimizar procesos clave como la organización, transferencia, consulta y conservación de documentos. Esta revolución tecnológica ha garantizado un acceso más eficiente y ágil a la información, convirtiéndola en un derecho fundamental que potencia la toma de decisiones informadas.

Es fundamental resaltar la transición hacia una gestión documental más inteligente y automatizada, que asegura que los recursos informativos estén disponibles en cualquier momento y lugar, y que sean útiles en el momento preciso. La Ciencia de Datos no solo ha mejorado las operaciones en los procesos de Gestión Documental, liderados por archivos públicos y privados, sino que también ha transformado el conocimiento en sabiduría práctica, impulsando el progreso de personas, empresas y organizaciones.

Recomendaciones

Optimizar almacenamiento: Utilizar soluciones de almacenamiento en la nube y bases de datos no relacionales para gestionar grandes volúmenes de datos de manera eficiente y segura, y así evitar costos a la hora de construir bodegas e instalación de mobiliarios para almacenar archivos en soporte palpable.

Mejorar la seguridad: Implementar medidas robustas de seguridad y encriptación para proteger los datos sensibles y garantizar el cumplimiento de normativas internacionales; para evitar ser víctima de técnicas o ataques cibernéticos, que suscitan pérdida de la credibilidad en las personas y empresas.

Capacitar al Talento Humano: Ofrecer formación continua a los empleados o servidores en el uso de nuevas tecnologías y herramientas de análisis de datos para maximizar su potencial y adaptabilidad.

Fomentar la toma de decisiones basada en datos (Data-Driven): Promover una cultura organizacional que valore el análisis de datos para la toma de decisiones informadas, mejorando así la eficiencia operativa y estratégica.

Referencias Bibliográficas

AGN - Colombia. (2004). *Memorias Seminario Archivos Policiales, Inspecciones de Policía.*

Recuperado el 13 de diciembre de 2024, de

https://biblioteca.archivogeneral.gov.co/pmb/opac_css/index.php?lvl=notice_display&id=21009

AGN - Colombia. (2016). *EXPERIENCIAS SIGNIFICATIVAS EN LA ADMINISTRACIÓN DE*

ARCHIVOS Y LA GESTIÓN DOCUMENTAL EN COLOMBIA: Policía Nacional de

Colombia. Recuperado el 2 de enero de 2025, de

<https://observatorioagn.archivogeneral.gov.co/wp-content/uploads/2020/08/EXPERIENCIAS-SIGNIFICATIVAS-POLICIA-NACIONAL-19-03-2017.pdf>

AGN - Colombia. (s.f.). *Formato Unico de Inventario Documental.* Recuperado el 18 de

diciembre de 2024, de Archivo General de la Nación - Colombia:

<https://www.archivogeneral.gov.co/inventario-documental>

arroba system. (30 de junio de 2022). Captura pantalla aplicativo Google Drive. Recuperado el 7

de enero de 2025, de <https://arrobasystem.com/blogs/blog/ve-facilmente-las-ubicaciones-de-los-archivos-en-google-drive>

Baratz. (7 de octubre de 2021). Recuperado el 13 de diciembre de 2024, de

<https://www.comunidadbaratz.com/blog/la-archivistica-a-lo-largo-de-la-historia/>

Bazán Gil, V. (2023). *Artificial intelligence applications in media archives.*

doi:<https://doi.org/10.3145/epi.2023.sep.17>

- Borras, G. J. (2021). *LA PLANIFICACIÓN ESTRATÉGICA DEL SERVICIO DE ARCHIVO Y GESTIÓN DOCUMENTAL: Qué es y cómo se implementa*. Recuperado el 1 de enero de 2025, de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/184193?page=12>
- Carrillo, A. (2008). *e-Administración*. UOC. Recuperado el 17 de diciembre de 2024, de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/lc/unad/titulos/56487>
- Carvajal, C. C. (9 de enero de 2011). *INFORME EJECUTIVO GESTIÓN SECRETARÍA GENERAL AÑO 2010*. Recuperado el 3 de enero de 2025, de http://polired/Institucion/NivelAsesor/Secregeneral/_layouts/15/WopiFrame.aspx?source doc=/Institucion/NivelAsesor/Secregeneral/Documents/INFORME%20DE%20GESTI%C3%93N%20SEGEN%202010.pdf&action=default&DefaultItemOpen=1
- Castillo Romero, J. (2019). *Big data*. IC. Recuperado el 4 de diciembre de 2024, de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/lc/unad/titulos/124254>
- Celaya Luna, A. (2014). *Cloud: herramientas para trabajar en la nube*. Recuperado el 4 de diciembre de 2024, de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/56046?page=8>
- Cerrillo Martínez, A. (2018). *Datos masivos y datos abiertos para una gobernanza inteligente*. Recuperado el 7 de diciembre de 2024, de <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2018.sep.16>
- Chávez, F. (2018). *¿Cómo funciona Internet?* Ciudad Educativa. Recuperado el 4 de diciembre de 2024, de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/lc/unad/titulos/36728>

- Cruz Herradón, A. M. (2013). *Internet y correo electrónico*. (ECP, Editor) Recuperado el 13 de diciembre de 2024, de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/50718?page=12>
- Cuellar, F., Augusto, M., Felacio, V., Agustin J, & Diaz, R. (2023). *Modelo de Big Data para la gestión documental en el área de certificación Centro agroindustrial SENA*. Obtenido de https://ciencia.lasalle.edu.co/maest_gestion_informacion_documental/33/
- De Bustos, P. P. (2007). *Sistemas integrados y gestión documental. La experiencia en Telecinco*. Obtenido de <https://research-ebSCO-com.bibliotecavirtual.unad.edu.co/linkprocessor/plink?id=62d2ddf2-8b89-3c48-8d0a-327bd779c0d6>
- Díaz Gonzalez, E. (2022). *La Gestión Documental empresarial ante el desafío de la transición digital*. Recuperado el 18 de diciembre de 2024, de <https://edicionsedic.es/clip/article/view/69/59>
- Díaz, G. E. (2022). *La Gestión Documental empresarial ante el desafío de la transición digital*. Obtenido de <https://edicionsedic.es/clip/article/view/69>
- Dominguez , L. (2021). *Análisis Del Modelo De Industria 4.0 Para La Transformación Tecnológica en El Sector Empresarial*. Recuperado el 1 de enero de 2025, de <https://research-ebSCO-com.bibliotecavirtual.unad.edu.co/c/qcagk4/viewer/pdf/3w5hr4a7zj>
- Escobar , B. M., Mercado, P. M., & Rodriguez , L. R. (2020). *Beneficios ofrecidos por la Gestión del Big Data en la era de la digitalización*. Obtenido de <https://revistas.usantotomas.edu.co/index.php/signos>

- García , V. E. (2016). *¿Cómo generar valor en las empresas?* (U. d. Pacífico, Ed.) Recuperado el 1 de enero de 2025, de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/173385?page=1>
- García Morales, E. (2013). *Gestión de documentos en la e-administración*. UOC. Recuperado el 17 de diciembre de 2024, de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/lc/unad/titulos/56710>
- GESDA. (2017). *Big Data y Gestión Documental*. Recuperado el 7 de diciembre de 2024, de <https://www.gedsa.es/big-data-y-gestion-documental>
- Godoy, J., López , I., & Casimillas, C. (2003). *Mini Manual No. 4 - TABLAS DE RETENCIÓN Y TRANSFERENCIAS DOCUMENTALES*. (I. N. Colombia, Ed.) Recuperado el 18 de diciembre de 2024, de https://www.archivogeneral.gov.co/sites/default/files/Estructura_Web/5_Consulte/Recursos/Publicacionees/Minimanual_TRD.pdf
- Gonzalo Edesa, V. (2018). *G Suite for Education: administración y configuración en centros educativos*. RA-MA. Recuperado el 6 de enero de 2025, de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/106251?page=1>
- Google. (s.f.). *Document AI*. Recuperado el 8 de enero de 2025, de <https://cloud.google.com/document-ai?hl=es-419>
- Gutierrez, E. (s.f.). *La Academia, Sesión 2: El Ciclo Vital del Documento y la confusión con las fases de archivo*. Recuperado el 18 de diciembre de 2024, de Nosonpapeles.com: <https://www.nosonpapeles.com/ciclo-vital-del-documento-fases-de-archivo/>

- Hegghammer, T. (22 de noviembre de 2021). *OCR with Tesseract, Amazon Textract, and Google Document AI: a benchmarking experiment*. Recuperado el 6 de enero de 2025, de <https://doi-org.bibliotecavirtual.unad.edu.co/10.1007/s42001-021-00149-1>
- Heredia Herrera, A. (1986). *Archivística general : teoría y práctica*. EXCMA. Recuperado el 2024 de diciembre de 4
- Huertas Bailén, A., Quintas Froute, N., & González Neira, A. (2024). *La Participación de la Audiencia en los Metadatos de TikTok*. Obtenido de <https://doi-org.bibliotecavirtual.unad.edu.co/10.58262/V32I78.7>
- Ibañez Montoya, J. (2008). *Los archivos: cómo construirlos*. Ediciones TREA. Recuperado el 17 de diciembre de 2024, de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/lc/unad/titulos/60540>
- Leal Buitrago, F. (13 de diciembre de 2015). Crisis en la Policía Nacional. *El Espectador*. Recuperado el 2 de enero de 2025, de <https://www.elespectador.com/colombia/mas-regiones/crisis-en-la-policia-nacional-article-605291/>
- Maldonado, S. (2022). *Analytics y Big Data: ciencia de los Datos aplicada al mundo de los negocios* (primera ed.). RIL editores. Recuperado el 16 de diciembre de 2024, de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/225562?>
- Mark, D. (2024). *Microsoft Excel*. Recuperado el 14 de diciembre de 2024, de <https://research-ebsco-com.bibliotecavirtual.unad.edu.co/c/qcagk4/viewer/html/o56gmkoe5>
- Marqués, M. P. (2015). *Big Data: Técnicas, Herramientas y aplicaciones*. Alfaomega.
- Marta. (2020). *Caso de éxito gestión documental - Laiex*. Recuperado el 1 de enero de 2025, de <https://gdx-group.com/caso-de-exito-gestion-documental-laiex/>

Martínez Gutiérrez, R. (2023). *Reforma Para Una Administración De Datos*. (R. C. Públíc, Ed.)

Recuperado el 22 de diciembre de 2024, de <https://doi-org.bibliotecavirtual.unad.edu.co/10.58992/rcdp.i67.2023.4094>

Mena Mujica, M. (2005). *Gestión documental y organización de archivos*. Fenix. Recuperado el 16 de diciembre de 2024

Mendez, A. (29 de noviembre de 2024). Policía de Colombia pionera en la utilización de la inteligencia artificial, IA, y así se proyecta para el 2050 y los ciberdelitos. *El Tiempo*. Recuperado el 3 de enero de 2025, de <https://www.eltiempo.com/justicia/delitos/policia-de-colombia-pionera-en-la-utilizacion-de-la-inteligencia-artificial-ia-y-asi-se-proyecta-para-el-2050-y-los-ciberdelitos-3404594>

Miroslav, F. (2024). *Google - Early history*. Recuperado el 6 de diciembre de 2024, de <https://research-ebSCO-com.bibliotecavirtual.unad.edu.co/c/qcagk4/viewer/html/oz6n4d6kvr>

Mojapelo, M., & Marutha, N. (2023). *Framework to Infuse Data Science in the Archives and Records Management Curriculum in South Africa*. Recuperado el 18 de diciembre de 2024, de <https://research-ebSCO-com.bibliotecavirtual.unad.edu.co/c/5y465o/viewer/pdf/5guen35yaf?route=details>

Mundet, J. R. (2012). *Archivística : gestión de documentos y administración de archivos*. Temis.

Muñoz M., J. (2009). *Historia de la computación y los documentos informáticos*. Recuperado el 4 de diciembre de 2024, de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/lc/unad/titulos/28595>

Policía Nacional. (s.f.). Recuperado el 3 de enero de 2025, de <https://www.policia.gov.co/quienes-somos>

- Policía Nacional. (2023). *El Plan Estratégico Institucional 2023 – 2026*. Recuperado el 3 de enero de 2024, de https://policia.edu.co/wp-content/uploads/2024/02/cartilla_bolsillo_pei_2023_2026.pdf
- Policía Nacional de Colombia. (2021). acceso Gestor de Documentos Policiales - GEPOL. Recuperado el 3 de enero de 2025, de <https://gepol.policia.gov.co>
- Policía Nacional de Colombia. (2021). Tablero de estadística Gestor de Documentos Policiales - GEPOL. Recuperado el 3 de enero de 2025, de <https://gepol.policia.gov.co>
- Ríos Hilario, A., & Fraile Sastre, A. (2020). *AL GRADO DE CONOCIMIENTO Y APLICACIÓN DE BIG DATA EN LAS BIBLIOTECAS UNIVERSITARIAS ESPAÑOLAS*. Recuperado el 19 de diciembre de 2024, de <https://revistas.um.es/analesdoc/article/view/390931/277421>
- Rodriguez de Sepulveda Mailillo, D. (2016). *Aplicaciones Google*. RA-MA. Recuperado el 6 de enero de 2025, de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/lc/unad/titulos/106397>
- Russo, P. (2013). *Gestión documental en las organizaciones*. (UOC, Ed.) Recuperado el 14 de diciembre de 2024, de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/56515?page=17>
- Sainz Gonzalez, M. (2023). *Más allá de los datos: transcodificación para visualizar la información*. Recuperado el 14 de diciembre de 2024, de <https://research-ebSCO-com.bibliotecavirtual.unad.edu.co/c/qcagk4/viewer/pdf/pwh7qh3avb?route=details>
- Sanabria, M., Gauthier Umanha, V., Méndez Rome, R. A., & Rivera Virgíez, L. (2022). *Transformación digital en las organizaciones*. Universidad del Rosario. Recuperado el 1 de enero de 2025, de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/219878?page=1>

Senado Colombia. (20 de julio de 1991). Obtenido de

http://www.secretariassenado.gov.co/senado/basedoc/constitucion_politica_1991.html

Téllez Carvajal, E. (2020). *Análisis documental sobre el tema del big data y su impacto en los derechos humanos*. (D. PUCP, Ed.) Recuperado el 22 de diciembre de 2024, de

<https://doi-org.bibliotecavirtual.unad.edu.co/10.18800/derechopucp.202001.006>

Toro López, F. (2023). *Ciencia de datos con Python*. ECOE. doi:9789585033566

Tukey, J. (1962). *The Future of Data Analysis*. Obtenido de

<https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-33/issue-1/The-Future-of-Data-Analysis/10.1214/aoms/1177704711.full>

Tukey, J. (1977). *Exploratory data analysis*. Recuperado el 13 de diciembre de 2024

Villar, A. (2021). *Acceso a la información pública y gestión documental: la experiencia uruguaya*. Recuperado el 18 de diciembre de 2024, de <https://research-ebSCO-com.bibliotecavirtual.unad.edu.co/c/5y465o/viewer/pdf/p2ufqgmbwj?route=details>

Viloria Ortín, M. (2023). *Inteligencia artificial y gestión documental en la gestión pública. Caso Perú*. Recuperado el 19 de diciembre de 2024, de

<https://revistauls.lasalle.edu.co/article/view/1836>

Vivas, E., Solanilla, L., & Alberch, R. (2013). *Digitalización del patrimonio: archivos,*

bibliotecas y museos en la red. UOC. Recuperado el 17 de diciembre de 2024, de

<https://elibro-net.bibliotecavirtual.unad.edu.co/es/lc/unad/titulos/56410>

White, T. (2015). *Hadoop: The Definitive Guide* (Cuarta ed.). (O'Reilly, Ed.) Recuperado el 16 de diciembre de 2024

Xu, D. (2022). *An Analysis of Archive Digitization in the Context of Big Data. Mobile*

Information Systems. Recuperado el 19 de diciembre de 2024, de <https://research-ebSCO-com.bibliotecavirtual.unad.edu.co/c/5y465o/viewer/pdf/sdr2q4cpxz>