



Análisis documental para orientar un manual de gestión y conservación de expedientes académicos

Documentary analysis to guide a manual for the management and preservation of academic records

Claudia Cristina Arizala Quintero^{1*}, Magda Emperatriz Segura Montezuma², Juan Francisco Cázares Zuleta³, Michelle Nicole García Ramírez⁴

¹ Universidad Politécnica Salesiana, Maestría en Gestión de la Calidad de la Información y Documentación, Quito, Ecuador, carizalaq@est.ups.edu.ec

² Universidad Politécnica Salesiana, Maestría en Gestión de la Calidad de la Información y Documentación, Quito, Ecuador, mseguram@est.ups.edu.ec

³ Universidad Politécnica Salesiana, Maestría en Gestión de la Calidad de la Información y Documentación, Quito, Ecuador, jcazaresz@est.ups.edu.ec

⁴ Universidad Politécnica Salesiana, Maestría en Gestión de la Calidad de la Información y Documentación, Quito, Ecuador, mgarcia12@est.ups.edu.ec

*Autora de correspondencia: Claudia Cristina Arizala Quintero.

Resumen

La gestión de expedientes académicos requiere relacionar organización, conservación y acceso con las necesidades institucionales. Este estudio analiza un corpus documental para orientar un futuro manual técnico-administrativo destinado a la Unidad Educativa Fiscomisional María Auxiliadora de Esmeraldas. Se revisaron los informes de los talleres del Grupo 3 y se contrastaron los cinco CSV corregidos con el Excel definitivo de 26 asociaciones y con el tablero. El corpus comprende 15 documentos; el modelo reúne 17 marcos, 8 temas, 2 registros temporales y 26 asociaciones, sin claves huérfanas ni duplicados en el grano documento-marco-tema. El cotejo literal reproducible localizó 7 de las 26 evidencias en los pasajes documentales disponibles de corpus.csv. El archivo adjunto permite recalcular 44.790 caracteres en el campo texto y 3.277 caracteres al excluir el segmento «CONTEXTO DEL INFORME»; las cifras históricas de 199.426 caracteres, 2.967 tokens útiles y 1.217 términos distintos del Taller 1 no se presentan como resultados reproducidos porque los insumos originales de procesamiento no están disponibles. La contribución del trabajo es un procedimiento trazable para revisar asociaciones y explicitar los límites del modelo como insumo para un futuro manual de expedientes académicos; no se atribuyen beneficios institucionales no evaluados.

Palabras clave: expedientes académicos; gestión documental; conservación documental; minería de texto; modelo estrella; ISO 15489.

Abstract

Academic records management requires linking organization, preservation, and access to institutional needs. This study analyzes a documentary corpus to inform a future technical and administrative manual for the Unidad Educativa Fiscomisional María Auxiliadora in Esmeraldas, Ecuador. The Group 3 workshop reports were reviewed, and the five corrected CSV files were checked against the definitive Excel workbook with 26 associations and the dashboard. The corpus comprises 15 documents; the model contains 17 frameworks, 8 topics, 2 temporal records, and 26 associations, with no orphan keys or duplicates at the document–framework–topic grain. A reproducible literal comparison located 7 of the 26 evidence strings in the documentary passages available in corpus.csv. The attached file allows recalculation of 44,790 characters in the text field and 3,277 characters after excluding the “CONTEXTO DEL INFORME” segment; the historical Workshop 1 figures of 199,426 characters, 2,967 useful tokens, and 1,217 distinct terms are not presented as reproduced results because the original processing inputs are unavailable. The contribution is a traceable procedure for reviewing associations and stating the model’s limits as input for a future academic-records manual; no unevaluated institutional benefits are claimed.

Keywords: academic records; records management; documentary preservation; text mining; star schema; ISO 15489.

1 Introducción

Los expedientes académicos permiten reconstruir actuaciones y trayectorias educativas. Su gestión exige conservar la relación entre documentos, actividades y responsables, además de facilitar consultas autorizadas. El concepto de archivo escolar también depende del nivel educativo y del contexto institucional; por ello, las prácticas universitarias no se trasladan automáticamente a una escuela [1]. ISO 15489-1 proporciona una referencia conceptual para la creación, captura y gestión de documentos de archivo en diferentes organizaciones [2].

La literatura ecuatoriana ofrece antecedentes sobre organización de archivos de gestión y series documentales. Palma Solórzano y colaboradores examinan procesos de unidades productoras universitarias y relacionan sus resultados con capacitación y procedimientos normalizados [3]. Desde el contexto mexicano, Sosa del Angel y colaboradores muestran cómo la gestión documental se articula con áreas responsables y etapas del ciclo documental [4]. Ambos antecedentes ayudan a formular preguntas sobre responsabilidades y procedimientos, aunque sus resultados pertenecen a instituciones y jurisdicciones específicas.

La guía de Rosero Sánchez aborda la organización y el almacenamiento de expedientes estudiantiles electrónicos en una facultad universitaria [5]. Por su parte, Zúñiga Mendoza propone un diagnóstico integral aplicado a un archivo histórico [6]. Estos trabajos permiten distinguir dos momentos: conocer las condiciones de gestión y, sobre esa base, diseñar instrumentos. La revisión latinoamericana de Zambrano Plúa y colaboradores también destaca la necesidad de fortalecer los sistemas de gestión documental y considerar la legislación de cada país [7].

El problema que motiva este artículo es cómo convertir referencias dispersas en una base verificable para un manual técnico-administrativo de gestión y conservación de expedientes académicos. La Unidad Educativa Fiscomisional María Auxiliadora de Esmeraldas constituye el destino propuesto de esa aplicación. La necesidad de un manual fue planteada por el equipo como punto de partida; este análisis bibliográfico no acredita por sí mismo su ausencia ni caracteriza las prácticas actuales de la institución.

El objetivo fue analizar los marcos y la calidad de la evidencia registrada en un corpus de 15 documentos para orientar el futuro manual. La pregunta del estudio fue qué referencias y limitaciones emergen al transformar ese corpus en datos consultables. El aporte consiste en conectar lectura documental, depuración de menciones y organización analítica, evitando que un conteo de palabras se interprete como prueba de aplicación normativa.



2 Materiales y métodos

2.1 Documentos y datos

Se desarrolló un estudio documental, descriptivo y exploratorio a partir del corpus utilizado por el Grupo 3 en la materia Análisis, Explotación y Visualización de Datos. Se trabajó con una selección previa de 15 documentos sobre archivos escolares, expedientes académicos y gestión documental educativa. No se presenta como una revisión sistemática exhaustiva ni como una muestra representativa de todas las instituciones educativas. La unidad de análisis fue el documento; para la organización de evidencias, se consideró la asociación entre documento, marco y tema.

El Taller 1 caracterizó los textos y sus relaciones léxicas; el Taller 3 organizó las asociaciones documentales en un modelo estrella. Para esta revisión se utilizaron como fuentes de control los cinco CSV corregidos, el libro definitivo modelo_estrella_FINAL_26.xlsx, el tablero de 26 asociaciones y corpus.csv. El Excel definitivo contiene 26 filas de hechos y sustituye a la versión intermedia de 29 que permanecía disponible en el cierre anterior. El cotejo de evidencias se realizó contra corpus.csv, separando el pasaje documental del segmento añadido «CONTEXTO DEL INFORME». Esta comprobación corresponde al reenvío del artículo y no se atribuye retrospectivamente a la ejecución original de los talleres. De estas fuentes de control, el paquete de evidencias adjunto a este reenvío incluye corpus.csv, el cotejo derivado cotejo_literal_26.csv, los cinco CSV corregidos y modelo_estrella_FINAL_26.xlsx; estos dos últimos se localizaron en los materiales de trabajo del equipo docente para la recalificación del ejercicio y se verificaron de forma independiente, cifra por cifra, antes de incluirse (ver Nota_procedencia_evidencia_adicional.txt en el material suplementario).

El paquete corregido prevaleció sobre los informes y archivos intermedios. Se distinguieron la presencia de una asociación, la etiqueta binaria desarrolla y la suficiencia textual registrada. La integridad se comprobó mediante unicidad de claves, pertenencia de claves foráneas y ausencia de duplicados documento–marco–tema. Las etiquetas se contabilizaron sin asumir que acreditan desarrollo sustantivo. No se reconstruyeron búsquedas ni exclusiones sin registro; las 57 fuentes del ejemplo docente no constituyen una población inicial demostrada del Grupo 3.

2.2 Transformación de los datos

La secuencia fue documentos → datos → análisis → resultado: extracción y mapeo textual, organización de asociaciones, consultas y orientación preliminar del manual. El grano es una asociación documento–marco–tema. HECHOS_MENCION tiene clave id_mencion y referencias id_documento, id_marco, id_tema y anio hacia las dimensiones documento, marco, tema y tiempo; estas claves son únicas en sus catálogos. La separación entre hechos y dimensiones responde al enfoque dimensional [8]. Los cinco CSV se contrastaron con sus hojas de Excel; el tablero incorpora cuatro conjuntos y omite el catálogo temporal.

2.3 Aprendizaje automático y análisis textual

El informe del Taller 1 documenta nube de palabras, exclusión de palabras vacías, normalización, frecuencias, n-gramas, similitud de Jaccard y una red de coocurrencia con exploración de comunidades. Jaccard compara conjuntos mediante la intersección dividida por la unión [9]. El informe descarta interpretar las comunidades de la red por falta de discriminación suficiente. No se acreditan entrenamiento predictivo ni evaluación de un modelo de aprendizaje automático. Los resultados léxicos históricos no se presentan como reproducidos en este artículo: aunque se conservaron las cifras declaradas de 199.426 caracteres, 2.967 tokens útiles y 1.217 términos distintos, faltan los archivos de normalización y procesamiento que permitirían recalcularlas de manera independiente. El análisis reproducible de este reenvío se limita a las consultas descriptivas, los controles de consistencia y las estadísticas calculables directamente desde corpus.csv.

2.4 Procedimiento de reproducibilidad

Para reproducir el cotejo literal se recorren las 26 filas de HECHOS_MENCION.csv. Para cada id_documento se toma de corpus.csv únicamente el texto situado antes del delimitador exacto «|| CONTEXTO DEL



INFORME:». Tanto el pasaje como la evidencia se normalizan únicamente colapsando secuencias de espacios y saltos de línea; no se eliminan tildes, no se cambia entre mayúsculas y minúsculas y no se aplican equivalencias semánticas. Una evidencia se considera localizada solo cuando la cadena completa aparece como subcadena del pasaje documental normalizado. Con este criterio se obtienen 7 coincidencias: mencion_0002, mencion_0003, mencion_0004, mencion_0008, mencion_0009, mencion_0015 y mencion_0017; las otras 19 no presentan coincidencia literal en la copia disponible. Para las estadísticas textuales reproducibles se cuenta la longitud de cada valor del campo texto de corpus.csv: 44.790 caracteres en total, promedio 2.986 y mediana 2.950. Al usar solo el segmento anterior al delimitador, quedan 3.277 caracteres documentales, promedio 218,47 y mediana 171. El archivo suplementario cotejo_literal_26.csv conserva el resultado fila por fila y el documento Procedimiento_reproducibilidad_Grupo3.docx describe los pasos de verificación. Para las coincidencias señaladas en la sección 3 como acompañadas de puntos suspensivos, se aplicó un criterio de solape: se admite que el punto final del fragmento de evidencia coincida con el primer punto de una elipsis de tres puntos inmediatamente contigua, ya sea que la elipsis preceda o siga al fragmento; sin este criterio de solape, únicamente cuatro de los siete casos muestran la elipsis estrictamente exterior al fragmento.

3 Resultados

La Tabla 1 presenta las cantidades declaradas para el modelo corregido, con verificación directa parcial en este reenvío (ver nota de fuente de la tabla). Las 26 menciones son asociaciones almacenadas, no documentos ni normas distintas. Solo 14 de los 17 marcos y 7 de los 8 temas intervienen en hechos. No hay claves huérfanas, duplicados del grano ni celdas vacías. Sin embargo, «no declarado» representa información faltante: aparece en el año de 14 documentos y en la fuente de derivación de los 8 temas. La dimensión temporal contiene 2018 y «no declarado»; no permite establecer un periodo representativo.

Tabla 1: Estructura verificada del modelo corregido

Componente	Cantidad	Unidad
Corpus documental	15	Documentos
Catálogo de marcos	17	Marcos
Catálogo temático	8	Temas
Dimensión temporal	2	Registros; incluye «no declarado»
Tabla de hechos	26	Menciones
Archivos del modelo	5	CSV

Fuente: cinco CSV corregidos y hojas equivalentes de modelo_estrella_FINAL_26.xlsx. El libro definitivo contiene 26 filas de hechos y sustituye al Excel intermedio de 29. Se cuentan filas de datos, excluyendo encabezados y filas finales vacías. Estos archivos fuente están incluidos en el paquete de evidencias adjunto a este reenvío. La verificación directa realizada para este reenvío recalculó, a partir de los cinco CSV, el catálogo completo (17 marcos, 8 temas y 2 registros temporales) y confirmó que coincide exactamente con las 26 menciones, 14 marcos y 7 temas activos que ya reproducía cotejo_literal_26.csv.

ISO 15489 tiene la mayor cobertura del registro: 7 de 15 documentos. No se establece una categoría de transversalidad, pues falta un umbral operativo documentado. El informe anterior consignaba 31 asociaciones. La versión intermedia retiró cuatro pares bibliográficos e incorporó dos repeticiones ($31 - 4 + 2 = 29$). La versión definitiva eliminó esas dos repeticiones y el registro de DIRKS procedente de un índice ($29 - 2 - 1 =$



26). Esta secuencia explica la depuración sin mezclar los resultados anteriores con los definitivos. La versión de 31 asociaciones se localizó y auditó de forma independiente en los registros institucionales del propio Taller 3A del equipo (31 filas, con el mismo catálogo de 17 marcos y los mismos 15 documentos), lo que corrobora el punto de partida declarado; no se incluye en el paquete de evidencias de este reenvío por tratarse de una entrega de curso previa a la depuración, no de material probatorio del resultado final. La versión intermedia de 29 sigue sin un archivo propio localizado y se mantiene como un paso declarado por el equipo, no auditado de forma independiente en esta revisión.

Tabla 2: Cobertura y condiciones registradas en el paquete corregido

Indicador	Resultado verificado en los registros
Documentos con ISO 15489	7 de 15 documentos
Menciones ISO con desarrolla=1	12 de 15 menciones ISO
Menciones supranacionales con desarrolla=1	4 de 6 menciones supranacionales
Documentos sin mención registrada	6 documentos
Documentos señalados por texto insuficiente	8 documentos
Leyes nacionales con mención	2; Costa Rica y Argentina
Marcos ecuatorianos registrados	0 en el catálogo del modelo

Fuente: cruces de HECHOS_MENCION.csv con DIM_DOCUMENTO.csv y DIM_MARCO.csv, comprobados en Excel. «Desarrolla» es una etiqueta registrada en el archivo, no una validación independiente de contenido. Estos archivos de cruce están incluidos en el paquete de evidencias adjunto a este reenvío. La verificación directa realizada para este reenvío recalculó, a partir de ellos, los 20 de 26 con desarrolla=1, los 12 de 15 ISO y los 4 de 6 supranacionales, los 14 documentos con año no declarado y los 8 documentos con texto insuficiente; todos coinciden exactamente con lo reportado en esta tabla.

Hay 20 asociaciones con desarrolla=1: 12 de las 15 clasificadas como ISO y 4 de las 6 supranacionales, además de otras familias. Son etiquetas sobre menciones y no expresan cumplimiento institucional. Nueve documentos registran al menos una mención y seis no registran ninguna: doc_05, doc_06, doc_08, doc_11, doc_14 y doc_15. Estos seis están entre los ocho etiquetados con texto insuficiente; los otros dos son doc_07 y doc_13. «Texto insuficiente» también se trata como un valor registrado, no como una evaluación semántica independiente.

El cotejo literal reproducido con el procedimiento descrito en 2.4 localizó 7 de las 26 evidencias en el pasaje documental correspondiente de corpus.csv: mencion_0002, mencion_0003, mencion_0004, mencion_0008, mencion_0009, mencion_0015 y mencion_0017. Las otras 19 no presentan coincidencia literal en el archivo disponible. Esto no demuestra que las asociaciones sean falsas, pero impide validarlas literalmente con esta copia y limita su trazabilidad. Cinco de las siete coincidencias se encuentran en pasajes que, en corpus.csv, contienen puntos suspensivos alrededor de otros fragmentos, bajo el criterio de solape definido en la sección 2.4; por ello, el contexto completo debe contrastarse con los originales antes de usar esas asociaciones como fundamento del futuro manual.

Las estadísticas reproducibles dependen del campo y del recorte aplicado. corpus.csv contiene 44.790 caracteres en el campo texto, incluidos los contextos añadidos; el promedio es 2.986 y la mediana 2.950. Al excluir «CONTEXTO DEL INFORME» quedan 3.277 caracteres del pasaje documental, con promedio 218,47



y mediana 171, coincidentes con la dimensión documental. El máximo documental es doc_12, con 995 caracteres; cinco registros carecen de pasaje documental: doc_06, doc_08, doc_11, doc_14 y doc_15. Las cifras históricas de 199.426 caracteres, 2.967 tokens útiles y 1.217 términos distintos corresponden al Taller 1, pero no son reproducibles con los archivos actualmente adjuntos; por ello no se utilizan como resultados recalculados de este artículo. El propio segmento «CONTEXTO DEL INFORME» de corpus.csv incluye, como texto narrativo, un desglose por documento (por ejemplo, «doc_01: 17.810» caracteres) que no coincide con los caracteres reales del campo texto de ese mismo documento en corpus.csv (3.693 caracteres); esta discrepancia interna refuerza la decisión de excluir dicho segmento del análisis reproducible.



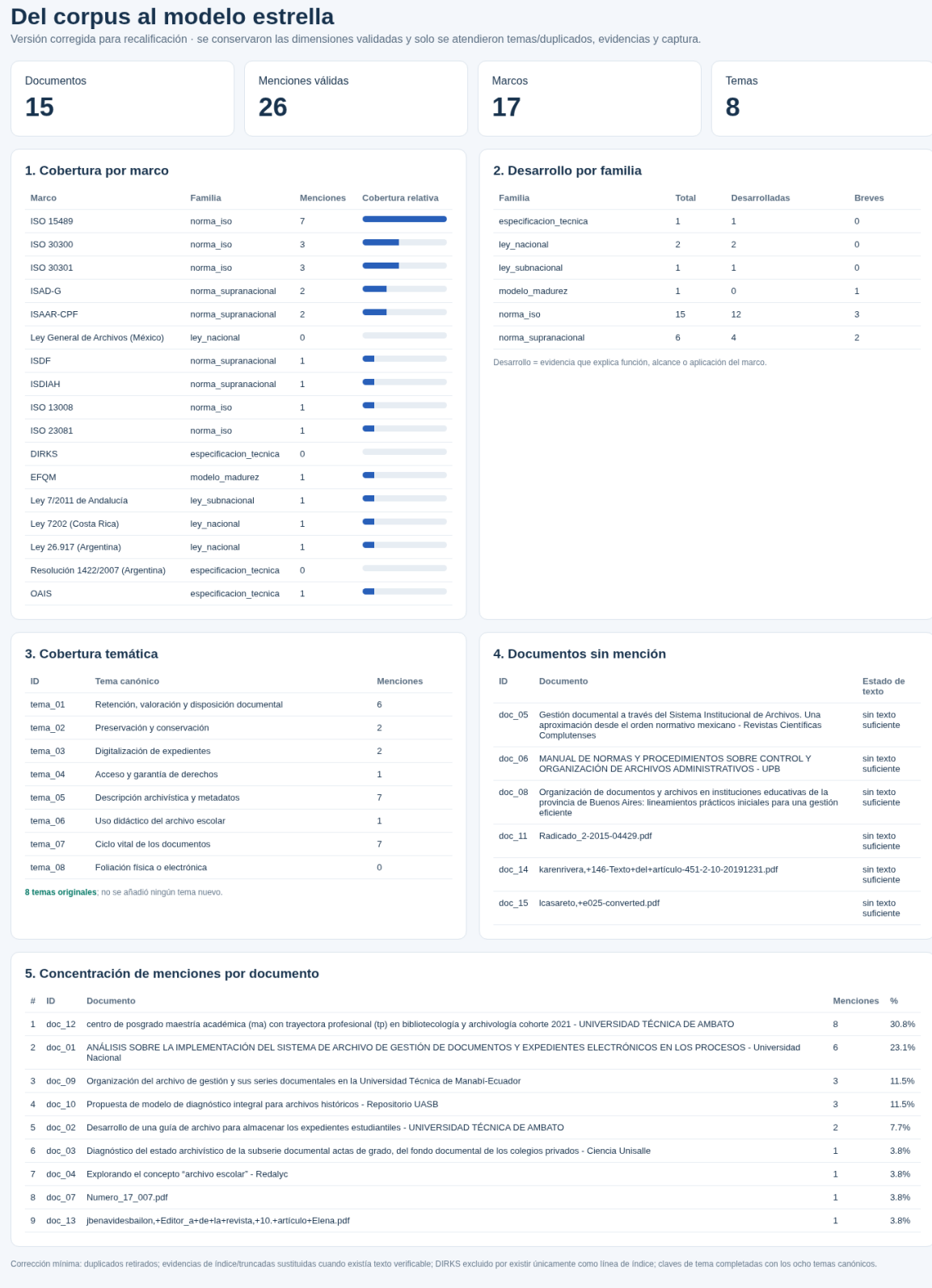


Figura 1: Captura del tablero. El equipo declaró como fuente un archivo llamado Grupo3_Recalificacion_100_VERIFICADO.zip que no se localizó con ese nombre exacto; sin embargo, se localizó y verificó por hash MD5 un archivo de imagen idéntico, byte a byte, a esta misma captura, dentro de los materiales de trabajo de la recalificación del equipo (incluido como captura_tablero_Grupo3.png en el paquete de evidencias de este reenvío; ver Nota_procedencia_evidencia_adicional.txt). Sus 26 «menciones válidas» son registros conservados y coinciden exactamente con los recalculados de forma independiente en las Tablas 1 y 2; esa coincidencia no acredita la validación semántica completa de las evidencias.



La Figura 1 conserva la captura original del tablero. Sus indicadores principales coinciden con los CSV; la concordancia numérica no resuelve las limitaciones textuales. La Figura 2 ilustra el entorno de aplicación previsto, sin representar instalaciones reales ni una intervención ejecutada.

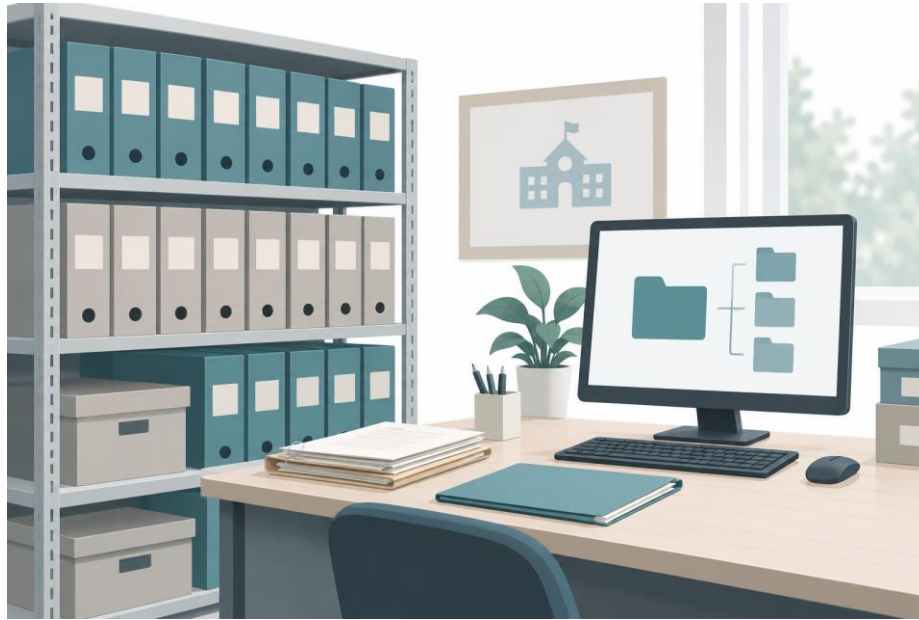


Figura 2: Representación conceptual de un archivo escolar con documentos físicos y recursos digitales. Ilustración generada con inteligencia artificial; no corresponde a la institución ni constituye evidencia empírica.

Capa complementaria: verificación neutrosófica de la cobertura documental (T, I, F)

Sin introducir cifras nuevas, los resultados ya reportados en las Tablas 1 y 2 admiten una lectura complementaria mediante la tripleta neutrosófica (T, I, F), que separa explícitamente lo que el corpus permite afirmar como desarrollado, lo que no permite determinar por insuficiencia de evidencia, y lo que permite afirmar como ausente pese a ser razonablemente esperado. Esta capa reagrupa los indicadores existentes bajo una notación que distingue indeterminación genuína de falsedad, en vez de forzar ambas categorías a un mismo valor por defecto.

T (verdad documental) se define aquí como la proporción de asociaciones documento–marco–tema con la etiqueta desarrolla=1 registrada en el archivo (Tabla 2): 20 de las 26 menciones registradas, $T \approx 0,77$. Como ya indica la nota de fuente de la Tabla 2, «desarrolla» es una etiqueta registrada, no una validación independiente de contenido ni una medida de cumplimiento institucional; T hereda esa misma limitación y se lee como la proporción de menciones etiquetadas, no como evidencia comprobada de desarrollo sustantivo. La cobertura no es uniforme entre familias normativas: 12 de 15 menciones de ISO 15489 ($T_{\text{ISO}} \approx 0,80$) y 4 de 6 menciones supranacionales ($T_{\text{supra}} \approx 0,67$).

I (indeterminación) no se calcula como el complemento aritmético de T, sino de forma independiente y en un grano distinto (documentos, no menciones), como la proporción de documentos del corpus para los que el propio modelo declara «sin texto suficiente» (Tabla 2): 8 de los 15 documentos, $I \approx 0,53$. Que T e I se definan en niveles de agregación distintos —menciones y documentos, respectivamente— es una decisión metodológica explícita, no un error de cómputo: cada indicador reutiliza el nivel en el que la Tabla 2 ya reporta el dato correspondiente, y ninguno de los dos se presenta como numéricamente comparable con el otro. En estos documentos no es posible determinar, con la evidencia disponible, si los marcos asociados están o no desarrollados; tratar esta indeterminación como si fuera falsedad (asumiendo $F=0$ por defecto ante la ausencia de texto) atribuiría al corpus una certeza que este no ofrece.

F (falsedad documental) se ancla en el hallazgo del corpus donde la ausencia está verificada, y no simplemente indeterminada: de los 17 marcos catalogados (Tabla 1), ninguno es de origen ecuatoriano, pese a que el manual futuro que este estudio busca orientar está destinado a una institución educativa ecuatoriana. Este hallazgo no es una insuficiencia de texto —los 17 marcos están íntegramente catalogados—, sino una ausencia de cobertura local que el propio corpus permite constatar. La comprobación de integridad referencial (ausencia de claves huérfanas en el grano documento–marco–tema, ya reportada en la Tabla 1) es un control de calidad de los datos, no una medida de falsedad documental, y por eso no se cuenta como componente de F.

Esta separación reproduce, a escala documental, una distinción epistemológica más general entre «no consta» —indeterminación genuina, aquí los ocho documentos con texto insuficiente— y «consta que no» —falsedad verificada, aquí la ausencia constatada de marcos de origen ecuatoriano—, evitando colapsar ambas categorías en un mismo valor por defecto al evaluar cobertura documental.

4 Discusión

La presencia de ISO 15489 permite proponerla como punto de partida para organizar criterios del futuro manual. La literatura latinoamericana también la utiliza para examinar necesidades de gestión documental [7]. Sin embargo, su frecuencia en el corpus no establece una jerarquía jurídica. La selección de procedimientos debe combinar pertinencia archivística, funciones escolares y requisitos aplicables, con una revisión contextual independiente del número de menciones.

No hay marcos ecuatorianos registrados en el catálogo; las dos leyes nacionales con hechos corresponden a Costa Rica y Argentina. El resultado justifica revisar la cobertura y completar la búsqueda local, sin afirmar ausencia normativa en las obras completas ni un vacío legal. Los antecedentes de México y Ecuador pertenecen a marcos organizativos propios [3], [4]; las leyes extranjeras sirven como referencias comparativas, no como obligaciones para la unidad educativa.

Como derivación propositiva, los resultados pueden orientar preguntas y criterios para un futuro manual: identificación del expediente, incorporación de documentos, clasificación por series, localización, consulta, seguimiento de préstamos, responsables y registros de control. La conservación y la disposición requerirían criterios institucionales y normativos verificados antes de fijar plazos. Esta derivación se limita a organizar trazabilidad entre fuentes, asociaciones y posibles componentes del manual; el estudio no evaluó implantación, reducción de tiempos, disminución de pérdidas documentales, mejora del acceso ni otros beneficios institucionales.

La vinculación entre soportes requiere datos descriptivos consistentes. ISO 23081-1 ofrece principios para los metadatos de gestión de documentos [10]. Para el caso propuesto, podría estudiarse un registro que relacione identificador, serie, fechas, ubicación, responsable y condiciones de acceso. Esta recomendación no presupone adquirir software ni digitalizar todos los expedientes: su propósito es mantener identificables las relaciones documentales y facilitar controles verificables.

El tamaño reducido del corpus, la heterogeneidad institucional y la extracción fragmentaria limitan la generalización. La base textual disponible tampoco permite reproducir las estadísticas léxicas históricas del Taller 1 ni localizar literalmente 19 de las 26 evidencias registradas. No se recuperaron las cadenas completas de búsqueda, los archivos de normalización ni pruebas de concordancia entre revisores. La consistencia referencial permite consultar el modelo y rastrear sus registros, pero no convierte sus etiquetas en validaciones semánticas independientes ni acredita causalidad, representatividad estadística o beneficios institucionales.

La siguiente etapa debería recuperar los textos completos o las ubicaciones originales de los 19 registros no localizados, documentar el criterio de suficiencia y contrastar cada asociación con su fuente. Después, sería necesario examinar los procesos de la unidad educativa y someter el futuro manual a revisión de las personas responsables. La contribución verificable de este trabajo es el esquema de trazabilidad —archivo de hechos,



dimensiones, corpus, cotejo literal y declaración de límites— que permite distinguir entre lo registrado, lo reproducido y lo que todavía requiere validación.

5 Conclusiones

El corpus de 15 documentos constituye una base preliminar para orientar el futuro manual de expedientes académicos. El paquete definitivo contiene 17 marcos, 8 temas y 26 asociaciones, con claves consistentes entre los cinco CSV y modelo_estrella_FINAL_26.xlsx; el Excel intermedio de 29 queda expresamente sustituido. ISO 15489 registra la mayor cobertura, con presencia en 7 documentos. Los 6 documentos sin menciones pertenecen al grupo de 8 etiquetados con texto insuficiente y permanecen dentro del corpus. El procedimiento reproducible localizó literalmente 7 de las 26 evidencias y documentó 19 no localizadas en la copia disponible. Asimismo, corpus.csv permite recalcular 44.790 caracteres del campo texto y 3.277 caracteres de pasajes documentales, mientras que las estadísticas léxicas históricas del Taller 1 no pueden reproducirse con los insumos actuales. El aporte se concentra en la trazabilidad de las asociaciones y en hacer explícitos los límites del modelo como orientación para un futuro manual; la aplicación institucional requiere recuperar y contrastar las fuentes y realizar un diagnóstico antes de atribuir beneficios o efectos a dicho manual.

Declaración de uso de inteligencia artificial

Este artículo se redactó y revisó con asistencia de ChatGPT (OpenAI), a partir de la información proporcionada por los autores, quienes revisaron el contenido y son responsables de él. Se conserva una ilustración conceptual generada con inteligencia artificial en la Figura 2. Los conteos estructurales, el cotejo literal 7/26 y las estadísticas directamente calculables desde corpus.csv se comprobaron en los archivos identificados; las cifras históricas no reproducibles y las limitaciones de evidencia se declaran expresamente.

Referencias

- [1] F. Frasson Martendal, “Explorando el concepto ‘archivo escolar’,” Información, cultura y sociedad, n.º 48, pp. 119–135, 2023. <https://doi.org/10.34096/ics.i48.11758>
- [2] International Organization for Standardization, ISO 15489-1:2016. Information and documentation — Records management — Part 1: Concepts and principles. Ginebra: ISO, 2016. <https://www.iso.org/standard/62542.html>
- [3] J. M. Palma Solórzano, L. M. Saltos Catagua, J. R. Beltrón Cedeño y K. V. Saldarriaga Villamil, “Organización del archivo de gestión y sus series documentales en la Universidad Técnica de Manabí-Ecuador,” ReHuSo, vol. 10, n.º 2, pp. 1–14, 2025. <https://doi.org/10.33936/rehuso.v10i2.7034>
- [4] C. O. Sosa del Angel, F. C. Caballero Rico, J. C. Guzmán García y C. Y. Perales Garza, “Gestión documental a través del Sistema Institucional de Archivos. Una aproximación desde el orden normativo mexicano,” Revista General de Información y Documentación, vol. 32, n.º 1, pp. 243–265, 2022. <https://doi.org/10.5209/rgid.82947>
- [5] S. del C. Rosero Sánchez, “Desarrollo de una guía de archivo para almacenar los expedientes estudiantiles digitalizados, optimizando los recursos en la Facultad Ciencias de la Salud,” trabajo de maestría, Universidad Técnica de Ambato, Ecuador, 2025. <https://repositorio.uta.edu.ec/server/api/core/bitstreams/15f1b186-b967-41ef-9ff9-ec4a7956600d/content>
- [6] R. D. Zúñiga Mendoza, “Propuesta de modelo de diagnóstico integral para archivos históricos: aplicación en el Archivo Histórico ‘Alfredo Pareja Diezcanseco’,” tesis de maestría, Universidad Andina Simón Bolívar, Sede Ecuador, Quito, 2018. <https://repositorio.uasb.edu.ec/handle/10644/6374>
- [7] I. E. Zambrano Plúa, E. M. Quindemil Torrijo y F. Rumbaut León, “Gestión documental en universidades: Una mirada desde Latinoamérica,” ReHuSo, vol. 6, n.º especial, pp. 108–119, 2021. <https://doi.org/10.33936/rehuso.v6iEspecial.3779>
- [8] R. Kimball y M. Ross, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3.ª ed. Wiley, 2013. <https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/books/data-warehouse-dw-toolkit/>
- [9] C. D. Manning, P. Raghavan y H. Schütze, Introduction to Information Retrieval. Cambridge University Press, 2008. <https://nlp.stanford.edu/IR-book/information-retrieval-book.html>



- [10] International Organization for Standardization, ISO 23081-1:2017. Information and documentation — Records management processes — Metadata for records — Part 1: Principles. Ginebra: ISO, 2017.
<https://www.iso.org/standard/73172.html>

