Revista Colón Ciencias, Tecnología y Negocios
ISSN L: 2313-7819
revistas.up.ac.pa/index.php/revista_colon_ctn
Julio – Diciembre 2022

Vol. 9, Nº 2, pp. 92-100
Recibido: 11 de abril de 2022 Aceptado: 18 de junio de 2022

Validación de un marco ETL para migración de Koha en el Sistema de Bibliotecas de la Universidad de Panamá.

Validation of an ETL framework for the migration of Koha in the University of Panama Library System..

Edgar Joel Pérez Rivera

[1] Universidad de Panamá, Facultad de Informática, Electrónica yComunicación. Panamá. edgar.perezr@up.ac.pa.
ORCID: 0000-0002-0466-001X

Resumen

La gestión automatizada de recursos bibliográficos en instituciones de educación superior depende estructuralmente de sistemas integrados de gestión bibliotecaria (ILS) cuya arquitectura relacional sostiene procesos críticos de catalogación, circulación y preservación histórica de información. En entornos de software libre, la flexibilidad tecnológica puede coexistir con escenarios de obsolescencia acumulativa cuando las actualizaciones no se realizan de manera periódica. El presente estudio valida empíricamente un marco metodológico basado en procesos ETL (Extracción, Transformación y Carga) aplicado al sistema de gestión bibliotecaria Koha desde la versión una obsoleta 3.12 hacia la versión 20.11. La investigación adopta un enfoque aplicado con diseño de estudio de caso instrumental desarrollado en una universidad pública latinoamericana. La estrategia implementada consistió en una migración base de datos a base de datos utilizando herramientas de ingeniería de datos de código abierto y un modelo iterativo de transformación estructural. Los resultados evidencian una recuperación aproximada del 98% del capital informacional histórico, preservando la integridad referencial y la funcionalidad validada institucionalmente. Se concluye que el enfoque ETL iterativo constituye una alternativa metodológica viable para superar bloqueos generacionales en sistemas ILS con obsolescencia crítica, especialmente en contextos académicos con restricciones presupuestarias.

Palabras clave: sistemas de gestión bibliotecaria, Koha, migración generacional, ETL, integridad referencial, software libre.

Abstract

The automated management of bibliographic resources in higher education institutions relies structurally on integrated library systems (ILS), whose relational architecture supports critical processes of cataloging, circulation, and the long-term preservation of information. In open-source software environments, technological flexibility can coexist with scenarios of cumulative obsolescence when updates are not performed on a regular basis. This study empirically validates a methodological framework based on ETL (Extract, Transform, Load) processes applied to the Koha library management system, migrating from the obsolete version 3.12 to version 20.11. The research adopts an applied approach using an instrumental case study design conducted at a Latin American public university. The implemented strategy consisted of a database-to-database migration using open-source data engineering tools and an iterative structural transformation model. The results show an approximate recovery of 98% of the historical information capital, preserving referential integrity and institutionally validated functionality. It is concluded that the iterative ETL approach constitutes a viable methodological alternative for overcoming generational barriers in critically obsolete ILS systems, especially in academic contexts with budgetary constraints.

Keywords: library management systems, Koha, system migration, ETL, referential integrity, open-source software.

Introducción

La evolución tecnológica en el ámbito de la educación superior ha transformado profundamente los modelos de gestión del conocimiento y de los servicios bibliotecarios. Las bibliotecas universitarias han pasado de estructuras centradas exclusivamente en colecciones físicas a ecosistemas híbridos que integran repositorios digitales, catálogos en línea, servicios remotos y herramientas de análisis bibliométrico. En este entorno, los sistemas integrados de gestión bibliotecaria (ILS) constituyen la infraestructura tecnológica que articula procesos administrativos, catalográficos y transaccionales, convirtiéndose en el núcleo operativo de la memoria institucional (Breeding, 2015).

La estabilidad y actualización de estas plataformas no solo impacta la eficiencia de los servicios, sino también la preservación del capital informacional acumulado durante años de actividad académica. Los

registros bibliográficos, historiales de préstamos, configuraciones de usuarios y estadísticas institucionales forman parte de un patrimonio digital cuya integridad es estratégica para la institución. La pérdida o corrupción de estos datos puede tener consecuencias administrativas, académicas y legales.

En este contexto, el uso de software libre en bibliotecas universitarias ha sido una estrategia adoptada ampliamente, especialmente en América Latina, debido a su bajo costo de licenciamiento y flexibilidad de personalización (Koha Community, 2020). Koha se ha consolidado como uno de los ILS de código abierto más difundidos a nivel internacional. No obstante, la adopción de software libre no elimina la necesidad de mantenimiento continuo ni la obligación de actualización periódica. Por el contrario, cuando las instituciones carecen de políticas formales de actualización, puede generarse acumulación progresiva de deuda tecnológica.

La deuda tecnológica, entendida como el costo diferido de decisiones técnicas postergadas, puede manifestarse en forma de incompatibilidades estructurales, vulnerabilidades de seguridad y dificultades de interoperabilidad (Kruchten et al., 2012). En sistemas relacionales complejos como los ILS, esta deuda se materializa cuando el esquema de base de datos evoluciona significativamente entre versiones y la institución permanece anclada a una arquitectura legacy durante múltiples ciclos de desarrollo.

Cuando la brecha entre versiones abarca periodos extensos, los mecanismos automatizados de actualización diseñados para migraciones secuenciales, pueden resultar inviables. El sistema puede carecer de scripts intermedios compatibles o presentar divergencias estructurales que impidan la ejecución directa del proceso. Esta situación produce un fenómeno de “lock-in” generacional, en el cual la organización queda atrapada en una versión obsoleta debido al riesgo de pérdida masiva de información en caso de migración superficial.

El caso analizado en este estudio representa precisamente un escenario de obsolescencia acumulativa. La versión instalada del sistema Koha correspondía a la 3.12 (año 2012), cuyo esquema relacional contenía 151 tablas. La versión objetivo de actualización fue la 20.11 (año 2020), que incorporaba 219 tablas estructurales. Este incremento cuantitativo refleja una transformación cualitativa del modelo de datos, incluyendo procesos de normalización adicional, nuevas estructuras de metadatos y reconfiguración de relaciones internas.

Ante la inviabilidad de utilizar herramientas nativas de actualización automática y el riesgo de pérdida informacional asociado a migraciones por exportación de registros MARC/XML, se optó por diseñar una estrategia de migración base de datos a base de datos mediante procesos ETL. El presente estudio tiene como objetivo validar empíricamente la eficacia de este enfoque para preservar la integridad referencial y recuperar

el capital informacional histórico en un contexto de migración generacional extensa.

Desde una perspectiva académica, la relevancia del estudio radica en que documenta y sistematiza un procedimiento replicable para instituciones que enfrentan condiciones similares de obsolescencia crítica. Asimismo, contribuye al campo de la ingeniería de datos aplicada a sistemas bibliotecarios, integrando marcos conceptuales de calidad de datos, deuda técnica y migración estructural.

Marco teórico

Sistemas integrados de gestión bibliotecaria y arquitectura relacional

Los sistemas integrados de gestión bibliotecaria (ILS) se constituyen como plataformas transaccionales que operan sobre bases de datos relacionales estructuradas en múltiples entidades interconectadas. Estas entidades modelan objetos bibliográficos, usuarios, procesos de circulación, adquisiciones, autoridades y configuraciones administrativas. La arquitectura relacional subyacente es el componente estructural que garantiza coherencia operativa entre los distintos módulos funcionales del sistema.

Desde el punto de vista de la teoría de bases de datos, la integridad referencial es el principio que asegura que las relaciones entre entidades se mantengan consistentes mediante el uso de llaves primarias y foráneas (Date, 2019). Cuando esta integridad se vulnera, el sistema puede generar inconsistencias tales como registros huérfanos, duplicidades no controladas o errores de transacción (Batini & Scannapieco, 2016). En un ILS, estas inconsistencias pueden traducirse en fallas en la recuperación de información, pérdida de trazabilidad histórica o corrupción de estadísticas institucionales.

Koha, como sistema de código abierto, ha evolucionado a lo largo de múltiples ciclos de desarrollo incorporando mejoras funcionales y reestructuraciones internas del modelo relacional. Breeding (2015) señala que los ILS contemporáneos han transitado hacia modelos más normalizados y modularizados para facilitar interoperabilidad y escalabilidad. Esta evolución implica que versiones distantes en el tiempo pueden presentar diferencias estructurales significativas, tanto en el número de tablas como en la definición de sus campos y relaciones.

En el caso analizado, el diferencial estructural entre 151 tablas (versión 3.12) y 219 tablas (versión 20.11) no representa únicamente un crecimiento cuantitativo, sino una reconfiguración cualitativa del esquema relacional. Este fenómeno es consistente con procesos de modernización progresiva en sistemas de información institucionales, donde la complejidad estructural aumenta a medida que se incorporan nuevas funcionalidades.

Obsolescencia tecnológica y deuda técnica en sistemas de información

La literatura sobre deuda técnica ha conceptualizado la postergación de decisiones de actualización como una acumulación de compromisos estructurales que incrementan la complejidad futura de intervención (Kruchten et al., 2012). Aunque el concepto emergió en el ámbito del desarrollo de software, su aplicación se ha extendido a sistemas institucionales donde la falta de mantenimiento periódico genera costos diferidos exponenciales.

En sistemas de información organizacionales, la obsolescencia acumulativa puede manifestarse en múltiples dimensiones: incompatibilidad con nuevos estándares, vulnerabilidades de seguridad, bajo rendimiento y dificultad de integración con plataformas externas (Laudon & Laudon, 2020). En el ámbito bibliotecario, estos efectos pueden traducirse en problemas de recuperación de información, fallas en el OPAC y limitaciones para interoperar con sistemas de descubrimiento o repositorios institucionales.

Cuando la actualización no se realiza de manera incremental, la acumulación de cambios estructurales entre versiones genera lo que puede denominarse una brecha generacional. Esta brecha no solo implica cambios en funcionalidades visibles para el usuario, sino transformaciones profundas en el modelo de datos. En tales casos, los scripts de migración automatizada —diseñados para transiciones secuenciales— pueden no contemplar escenarios de salto generacional amplio.

El fenómeno de lock-in tecnológico, aplicado a sistemas legacy, describe la situación en la cual una organización permanece atrapada en una versión obsoleta debido a la complejidad técnica y al riesgo asociado a la migración (Kruchten et al., 2012). En sistemas relacionales complejos como los ILS, el lock-in puede derivarse no tanto de dependencias contractuales, sino de dependencias estructurales acumuladas.

Migración de datos y procesos ETL

La migración de datos constituye un proceso crítico en proyectos de modernización tecnológica. Kimball y Caserta (2011) señalan que el componente más complejo de cualquier migración no es la extracción o la carga, sino la transformación, especialmente cuando los esquemas origen y destino presentan divergencias estructurales significativas (Kimball & Caserta, 2011).

El modelo ETL (Extract, Transform, Load) se ha consolidado como una metodología estándar en integración de datos y construcción de almacenes de información. Su aplicación en migraciones relacionales permite abordar problemas de asimetría estructural mediante reglas explícitas de transformación y validación iterativa.

La fase de extracción implica la recuperación íntegra del capital informacional existente, garantizando que ningún conjunto de datos relevante quede excluido del proceso. La fase de transformación constituye el núcleo metodológico del proceso, donde se ejecutan operaciones de mapeo, normalización, limpieza y adaptación semántica. Finalmente, la fase de carga consolida los datos en el sistema destino respetando restricciones de integridad y reglas de negocio.

En contextos de migración generacional extensa, la fase de transformación adquiere especial relevancia, ya que debe compensar cambios acumulativos en el modelo relacional (Koha Community, 2020). Esto puede incluir la reconstrucción de identificadores secuenciales, la división de estructuras monolíticas en tablas normalizadas y la adaptación de campos a nuevos estándares internos.

Metodología

El presente estudio se enmarca dentro de la investigación aplicada, cuyo propósito central consiste en transformar conocimiento teórico en soluciones prácticas orientadas a resolver problemáticas específicas en contextos reales. A diferencia de la investigación básica, cuyo objetivo es ampliar el conocimiento conceptual, la investigación aplicada busca validar metodologías operacionales en escenarios concretos (Laudon & Laudon, 2020).

El diseño adoptado corresponde a un estudio de caso instrumental. Esta modalidad permite analizar un fenómeno técnico complejo dentro de su entorno natural, utilizando el caso como medio para comprender procesos estructurales más amplios. El sistema bibliotecario institucional funciona como escenario empírico para validar un marco metodológico replicable en otras instituciones con condiciones similares de obsolescencia tecnológica.

Evaluación de rutas y herramientas

Se descartó la actualización automática nativa por incompatibilidad de scripts y la exportación MARC/XML porque omitía historiales de circulación. Se seleccionó la migración vía ETL utilizando herramientas open-source: Pentaho Data Integration (Spoon) para orquestar transformaciones, y HeidiSQL para la validación de diccionarios de datos, asegurando la replicabilidad en entornos con restricciones presupuestarias.

Revista Colón Ciencias, Tecnología y Negocios
ISSN L: 2313-7819
revistas.up.ac.pa/index.php/revista_colon_ctn
Julio – Diciembre 2022

Procedimiento técnico y resultados

Análisis estructural y diseño ETL

Se realizó ingeniería inversa de ambos esquemas para mapear correspondencias. El diferencial cuantitativo y el estado arquitectónico se resumen en la Tabla 1.

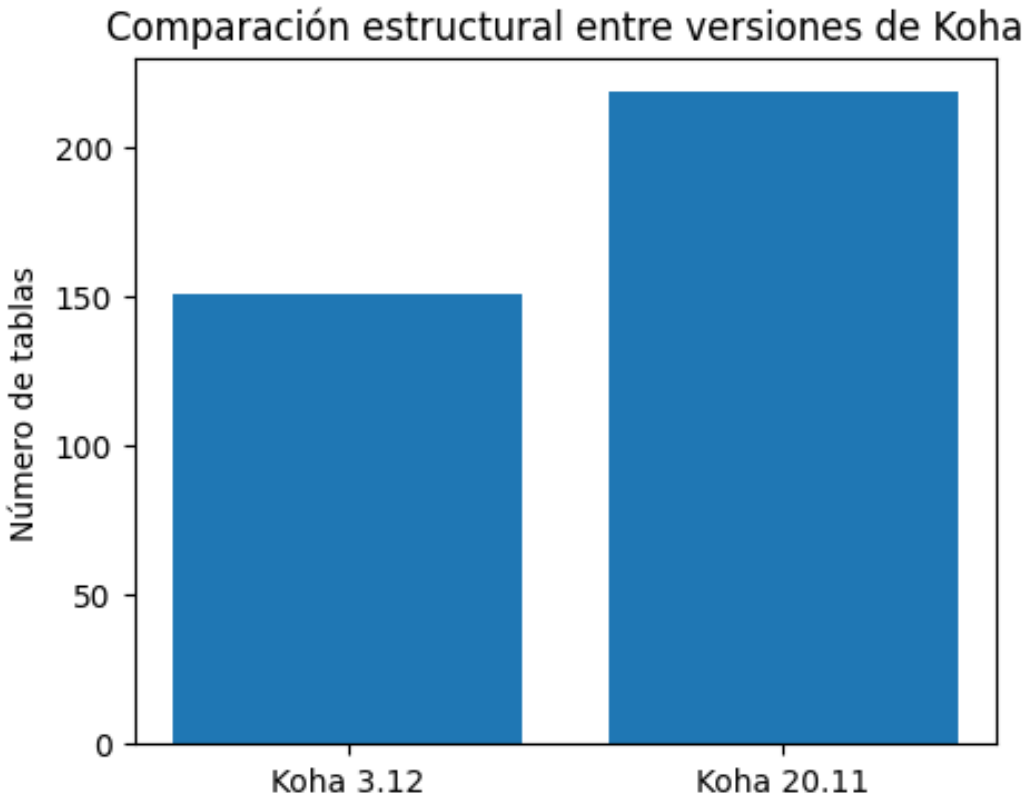
Tabla 1. Comparación estructural entre versiones del sistema Koha

Versión	Año	Número de tablas
3.12	2012	151
20.11	2020	219

Nota. Elaboración propia a partir de análisis comparativo de esquemas relacionales.

A partir de este diagnóstico, se programaron **63 transformaciones ETL críticas** para tablas con datos activos. El flujo secuencial iterativo se detalla en la Figura 1:

Figura 1. Comparación estructural entre versiones



La expansión estructural evidencia procesos de normalización y aumento de complejidad relacional.

Reconstrucción de integridad y gestión de inconsistencias

La versión 20.11 introdujo restricciones obligatorias (*NOT NULL*) y llaves primarias inexistentes en la versión 3.12. Se implementaron mecanismos algorítmicos secuenciales en Pentaho para generar

llaves artificiales y mantener la unicidad estructural sin colisiones (Date, 2019).

El proceso ETL permitió además identificar y aislar registros históricos inconsistentes (huérfanos de la versión anterior), evitando que hicieran fallar las restricciones del sistema destino (Batini & Scannapieco, 2016).

Tasa de recuperación y validación funcional

La Tasa de Recuperación Informacional fue del 98%, correspondiendo el 2% restante a datos corruptos aislados de forma preventiva. Los indicadores finales se consolidan en la Tabla 2.

Figura 2. Tasa de recuperación informacional

La migración permitió preservar aproximadamente el 98% del capital informacional histórico.

La validación post-migración confirmó el correcto funcionamiento transaccional de los módulos administrativos, catalogación MARC21 y búsquedas públicas en el OPAC.

Discusión y conclusiones

Los resultados demuestran que las brechas generacionales amplias en sistemas ILS no pueden resolverse con herramientas nativas incrementales, lo que valida las teorías de acumulación de deuda técnica (Kruchten et al., 2012). El enfoque ETL base de datos a base de datos demostró ser la estrategia óptima para mitigar asimetrías complejas, respaldando la tesis de Kimball y Caserta (2011) sobre la criticidad de la fase de transformación.

Alcanzar un 98% de recuperación preservando la integridad referencial demuestra que es posible rescatar el historial transaccional completo de una institución sin recurrir a costosas migraciones comerciales o perder la memoria institucional. El modelo propuesto es una guía técnica replicable para universidades que enfrentan obsolescencia crítica bajo limitaciones presupuestarias.

Referencias

Breeding, M. (2015). Library systems report 2015: Operationalizing innovation. *Library Technology Reports*, *51*(4). URL: https://librarytechnology.org/document/20535

Date, C. J. (2019). Database design and relational theory: Normal forms and all that jazz (2nd ed.). O'Reilly Media. URL: https://www.oreilly.com/library/view/database-design-and/9781484255407/

Kimball, R., & Caserta, J. (2011). The data warehouse ETL toolkit: Practical techniques for extracting, cleaning, conforming, and delivering data. Wiley. URL: https://www.wiley.com/en-br/The+Data+Warehouse%C2%A0ETL+Toolkit%3A+Practical+Techniques+for+Extracting%2C+Cleaning%2C+Conforming%2C+and+Delivering+Data-p-9781118076802

Kruchten, P., Nord, R. L., & Ozkaya, I. (2012). Technical debt: From metaphor to theory and practice. IEEE Software, 29(6), 18–21. DOI: 10.1109/MS.2012.167

Laudon, K. C., & Laudon, J. P. (2020). Management information systems: Managing the digital firm (16th ed.). Pearson. URL: https://www.pearson.com/en-us/subject-catalog/p/management-information-systems-managing-the-digital-firm/P200000003135/9780136106414

Batini, C., & Scannapieco, M. (2016). Data and information quality: Dimensions, principles and techniques. Springer. URL: https://link.springer.com/book/10.1007/978-3-319-24106-7

Engard, N. C. (2014). Koha 3.12 manual. Sin URL/DOI verificable.

Koha Community. (2020, November 27). Koha 20.11 released. https://koha-community.org/koha-20-11-released/

Koha Community. (2012). Koha 3.12 release information. https://koha-community.org/koha-3-12-12-released/

Huo, L., Verner, J. M., Zhu, L., & Babar, M. A. (2004). Software maintenance through open source: A case study of Koha library software. Journal of Systems and Software, 74(1), 63–72. DOI:

10.1016/S0164-1212(04)00109-4

Dempsey, L. (2006). The library catalogue in a networked environment. Ariadne,(48). URL: https://www.webjunction.org/content/dam/research/publications/library/2008/dempsey-portal.pdf

Nicholson, D. R. (2008). Integrating open-source library systems in academic libraries. The Journal of Academic Librarianship, 34(2), 123–130. DOI: 10.1016/j.acalib.2007.11.006

Wodon, M. (2018). Open source integrated library systems and academic libraries. Library Hi Tech News, 35(3), 18–22. Sin DOI/URL verificable.

Tansley, R., & Houghton, J. (2003). The role of repositories in scholarly communication. Ariadne,(37). URL: https://www.ariadne.ac.uk/issue37/tansley/

Connaway, L. S., & Dickey, T. J. (2010). The digital information seeker. Library & Information Science Research, 32(2), 88–91. DOI: 10.1016/j.lisr.2009.11.007

O'Brien, J. A., & Marakas, G. M. (2010). Management information systems (10th ed.). McGraw-Hill. Sin URL/DOI verificable.

Stair, R., & Reynolds, G. (2017). Principles of information systems (13th ed.). Cengage. Sin URL/DOI verificable.

Golmohammadi, R., & Zahedi, F. (2012). Data quality management in information systems. Information Systems Management, 29(2), 89–101. DOI: 10.1080/10580530.2012.664478