



# **UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

POSGRADO EN BIBLIOTECOLOGÍA Y ESTUDIOS DE LA INFORMACIÓN

FACULTAD DE FILOSOFÍA Y LETRAS

INSTITUTO DE INVESTIGACIONES BIBLIOTECOLÓGICAS Y DE LA INFORMACIÓN

**DISEÑO DE UN MODELO DE BIBLIOMINERÍA Y ESTRUCTURACIÓN DE UN DATA WAREHOUSE PARA IDENTIFICAR PATRONES DE CONOCIMIENTO SIGNIFICATIVO EN EL USO DE MATERIALES DOCUMENTALES**

**TESIS**

QUE PARA OPTAR POR EL GRADO DE  
DOCTORA EN BIBLIOTECOLOGÍA Y ESTUDIOS DE LA INFORMACIÓN

PRESENTA:

***GABRIELA MENDOZA GUILLÉN***

TUTOR PRINCIPAL:

*DR. SALVADOR GORBEA-PORTAL*

*INSTITUTO DE INVESTIGACIONES BIBLIOTECOLÓGICAS Y DE LA INFORMACIÓN*

MIEMBROS DEL COMITÉ TUTOR:

*DRA. MA. MARICELA PIÑA POZAS*

*DR. JUAN VOUTSSÁS MÁRQUEZ*

*INSTITUTO DE INVESTIGACIONES BIBLIOTECOLÓGICAS Y DE LA INFORMACIÓN*

**Ciudad Universitaria, CD. MX., Septiembre, 2022**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## **AGRADECIMIENTOS**

A mi tutor principal, Dr. Salvador Gorbea Portal, por brindarme su valioso saber, guía, apoyo y paciencia durante todos estos años.

A mis cotutores, Dr. Juan Voutssás Márquez y Dra. Maricela Piña Pozas, por compartir sus conocimientos y por todo el soporte otorgado.

A la Dra. Jane Russell Barnard por brindar valiosas observaciones en la etapa inicial de la investigación.

A la Dra. Lina Escalona Ríos, quien de manera comprensiva, firme y paciente me impulsó para finalmente llevar a cabo este proyecto.

A RM por sumar a la causa con su amplio conocimiento.

Al personal del SUBA UACH que colaboró en la investigación, en especial a Fabiola y a los hermanos Arenivas, quienes amablemente aportaron su experiencia en el sistema bibliotecario y en la biblioteca de la DES de Salud de la UACH, además de atender las solicitudes de información necesarias para la realización de esta tesis.

A amistades y colegas consultados para la obtención de información y el esclarecimiento de dudas referentes a la tesis, quienes generosamente compartieron su tiempo, ideas, conceptos y recursos.

A mis sinodales por todo el apoyo ofrecido.

Al personal de la Coordinación de Posgrado en Bibliotecología y Estudios de la Información, que de manera atenta ayudaron en la realización de los trámites administrativos.

A la Universidad Nacional Autónoma de México y al apoyo otorgado por el CONACyT.

A Paula, Paúl y a mi Lu, sin ustedes esto no habría sido posible, LAI.

# TABLA DE CONTENIDO

RESUMEN.....	X
ABSTRACT .....	XI
<b>INTRODUCCIÓN .....</b>	<b>XII</b>
<b>CAPÍTULO 1. EVALUACIÓN DE LA BIBLIOTECA UNIVERSITARIA MEDIANTE APLICACIONES DE LA CIENCIA DE DATOS .....</b>	<b>1</b>
1.1 BIBLIOTECAS UNIVERSITARIAS .....	2
1.2 ESTUDIOS MÉTRICOS DE LA INFORMACIÓN.....	5
1.3 CIENCIA DE DATOS .....	7
1.4.1 DESCUBRIMIENTO DE CONOCIMIENTO A PARTIR DE BASES DE DATOS .....	9
1.4.2 DATA WAREHOUSE .....	14
1.4.3 MINERÍA DE DATOS.....	19
1.4.4 HABILIDADES Y ROLES DE LOS PARTICIPANTES EN LA CIENCIA DE DATOS .....	23
1.5 APLICACIONES DE DATA WAREHOUSE Y MINERÍA DE DATOS EN BIBLIOTECAS .....	26
<b>CAPÍTULO 2. METODOLOGÍA .....</b>	<b>34</b>
2.1 UNIDAD DE ANÁLISIS Y OBSERVACIÓN .....	34
2.2 DISEÑO DEL DATA WAREHOUSE .....	49
2.2.1 ANÁLISIS DE REQUERIMIENTOS.....	50
2.2.2 ANÁLISIS DE LAS FUENTES DE DATOS.....	69
2.2.3 MODELO LÓGICO DEL DATA WAREHOUSE.....	94
2.2.4 INTEGRACIÓN DE LOS DATOS.....	105
2.3 DISEÑO Y APLICACIÓN DE LA BIBLIOMINERÍA .....	144
2.3.1 NOTA PREVIA: BIBLIOMINERÍA Y LA PRIVACIDAD DE LOS USUARIOS .....	144
2.3.2 PROGRAMA PARA REALIZAR EL EJERCICIO DE BIBLIOMINERÍA: PYTHON .....	147
2.3.3 APLICACIÓN DE UN PROCESO DE BIBLIOMINERÍA EN LA BDESSALUD .....	147
<b>CAPÍTULO 3. ANÁLISIS DE RESULTADOS .....</b>	<b>163</b>
3.1 RESULTADOS DE CONSULTAS HECHAS AL DATA WAREHOUSE DE LA BDESSALUD .....	163
3.2 RESULTADOS DEL EJERCICIO DE BIBLIOMINERÍA.....	187
3.3 MODELO DE BIBLIOMINERÍA PROPUESTO .....	191
<b>CAPÍTULO 4. DISCUSIÓN.....</b>	<b>192</b>

4.1	SOBRE LOS PLANTEAMIENTOS INICIALES DE LA TESIS.....	192
4.2	SOBRE LOS RESULTADOS Y APORTES DEL ESTUDIO.....	201
4.3	LIMITANTES DEL ESTUDIO.....	205
4.4	PROYECCIONES PARA EL ESTUDIO.....	208
<b><u>CAPÍTULO 5. CONCLUSIONES Y RECOMENDACIONES.....</u></b>		<b><u>213</u></b>
5.1	CONCLUSIONES.....	214
5.2	RECOMENDACIONES.....	218
<b><u>REFERENCIAS.....</u></b>		<b><u>220</u></b>
<b><u>ANEXO 1. RESUMEN DE LA MATRIZ DE MEDICIÓN Y EVALUACIÓN HOLÍSTICA DE NICHOLSON.....</u></b>		<b><u>230</u></b>
<b><u>ANEXO 2. PROVEEDORES DE SERVICIOS DE INFORMACIÓN ESPECIALIZADA PARA LA TOMA DE DECISIONES EN BIBLIOTECAS.....</u></b>		<b><u>231</u></b>
<b><u>ANEXO 3. NORMAS DE OPERACIÓN DEL SUBA-UACH.....</u></b>		<b><u>232</u></b>
<b><u>ANEXO 4. ENCUESTA SOBRE CONDUCTA INFORMATIVA DE PROFESIONALES DE LA INFORMACIÓN..</u></b>		<b><u>233</u></b>
<b><u>ANEXO 5. PARÁMETROS Y VARIABLES INCLUIDOS EN CADA CATEGORÍA DE REPORTES DEL LOGALEPH DEL SUBA-UACH.....</u></b>		<b><u>239</u></b>
<b><u>ANEXO 6. VARIABLES DE INTERÉS VS VARIABLES DISPONIBLES EN LA BDESSALUD.....</u></b>		<b><u>243</u></b>
<b><u>ANEXO 7. POLÍTICAS DE CARGA PARA EL LLENADO DEL DW BDESSALUD.....</u></b>		<b><u>251</u></b>
<b><u>ANEXO 8. POLÍTICAS Y ESTRATEGIAS DE ACTUALIZACIÓN DEL DW BDESSALUD.....</u></b>		<b><u>252</u></b>
<b><u>ANEXO 9. CATÁLOGO DE CÓDIGOS PARA VALORES DE LAS TABLAS ORACLE ALEPH DEL SUBA.....</u></b>		<b><u>253</u></b>
<b><u>ANEXO 10. POLÍTICA DE SEGURIDAD DE DATOS DEL SUBA-UACH.....</u></b>		<b><u>259</u></b>
<b><u>ANEXO 11. TEMÁTICAS PARA CONSIDERAR EN LA FORMACIÓN DE LOS BIBLIOTECARIOS DE DATOS.</u></b>		<b><u>260</u></b>
<b><u>ANEXO 12. IDENTIFICACIÓN DE INICIATIVAS DE APLICACIÓN DE METODOLOGÍAS DE LA CIENCIA DE DATOS EN BIBLIOTECAS MEXICANAS.....</u></b>		<b><u>261</u></b>

## Lista de siglas y abreviaturas utilizadas

**ACRL** *Association of College and Research Libraries* – Asociación de Bibliotecas Universitarias y de Investigación de Estados Unidos

**ACRR2** Reglas de Catalogación Angloamericanas, segunda edición (2004)

**ALA** *American Library Association*

**ALEPH** *Automated Library Expandable Program Hebreum* - Programa Hebreo Expandible para Automatización de Bibliotecas

**ANUIES** Asociación Nacional de Universidades e Instituciones de Educación Superior

**API** *Application Programming Interface* – Interfaz de programación de aplicaciones

**ARC** *Aleph Reporting Center*

**BDESSalud** Biblioteca de la Dependencia de Educación Superior de Salud del SUBA, UACH

**BI** *Business Intelligence* - Inteligencia empresarial o de negocios

**BVS** Biblioteca Virtual en Salud

**CACEI** Consejo de Acreditación de la Enseñanza de la Ingeniería, A. C.

**CCSUBA** Consejo Consultivo del Sistema Universitario de Bibliotecas Académicas

**CENETEC** Centro Nacional de Excelencia Tecnológica en Salud

**CIEES** Comités Interinstitucionales para Evaluación de la Educación Superior

**CIMAT** Centro de Investigaciones en Matemáticas, A. C.

**COMACE** Consejo Mexicano para la Acreditación de Enfermería, A. C.

**COMAEM** Consejo Mexicano para la Acreditación de la Educación Médica, A. C.

**CONACYT** Consejo Nacional de Ciencia y Tecnología

**CONCAPREN** Consejo Nacional para la Calidad de Programas Educativos en Nutriología, A. C.

**CONPAB-IES** Consejo Nacional para Asuntos Bibliotecarios de las Instituciones de Educación Superior

**CONRICyT** Consorcio Nacional de Recursos de Información Científica y Tecnológica

**COPAES** Consejo para la Acreditación de la Educación Superior, A. C.

**COUNTER** *Counting Online Usage of Networked Electronic Resources* - Contando el uso en línea de recursos electrónicos en red (Estándar)

**CSV** *Comma-Separated Values* - Valores separados por comas

**DAMA International** *Data Management Association* - Comunidad Global de Gestión de Datos

**DecS** Descriptores en Ciencias de la Salud

**DES de Salud** – Dependencia de Educación Superior de Salud (Incluye la Facultad de Medicina y Ciencias Biomédicas y la Facultad de Nutriología y Enfermería de la UACH)

**DES** Dependencia de Educación Superior

**DGTIC** Dirección General de Cómputo y de Tecnologías de Información y Comunicación

**DIKW** *Data Information Knowledge Wisdom* – Proceso de transformación del dato a información a conocimiento a sabiduría

**DOAJ** *Directory of Open Access Journals*

**DSI** Diseminación selectiva de información

**DSS** *Decision Support System* - Sistema de apoyo para la toma de decisiones

**DW** *Data Warehouse* - Repositorio o almacén de datos.

**EBL** *Evidence-based Librarianship* - Biblioteconomía basada en evidencias

**EGEL** Examen General para el Egreso de la Licenciatura

**EMI** Estudios Métricos de la Información

**ETL** *Extraction, transformation and loading* - Proceso de extracción, transformación y carga de los datos

**FEN** Facultad de Enfermería y Nutriología de la Universidad Autónoma de Chihuahua

**FMCB** Facultad de Medicina y Ciencias Biomédicas de la Universidad Autónoma de Chihuahua

**GPU** *Graphic Processing Unit* - Unidad de Procesamiento de Gráficos

**GUI** *Graphic User Interface* - Interfaz Gráfica de Usuario

**IA** Inteligencia Artificial (*AI*, por el anglicismo *Artificial Intelligence*)

**IES** Institución de Educación Superior

**IFLA** *International Federation of Library Associations and Institutions* - Federación Internacional de Asociaciones e Instituciones Bibliotecarias

**INEGI** Instituto Nacional de Estadística y Geografía

**IPN** Instituto Politécnico Nacional

**ISBN** *International Standard Book Number* - Código normalizado internacional para libros

**ISO** *International Organization for Standardization*

**ISSN** *International Standard Serial Number* – Código normalizado internacional que identifica todos los recursos continuos, independientemente de su formato (impreso o electrónico)

**JB** Jef@ de biblioteca del SUBA

**JU** Jef@ de unidad del CCSUBA

**KDD** *Knowledge Discovery in Databases* - Descubrimiento de conocimiento a partir de bases de datos

**LC** *Library of Congress* - Biblioteca del Congreso de los Estados Unidos

**MARC 21** *Machine-Readable Cataloging*

**MD** Minería de datos (en inglés *Data Mining*)

**MDoc** Material documental

**ML** *Machine Learning* - Aprendizaje automático

**MySQL** Servicio de sistemas de administración o gestión de bases de datos relacionales (**RDBMS**) de código abierto respaldado por Oracle y fundamentado en el lenguaje de consulta estructurada (**SQL**). Funciona bajo un modelo cliente-servidor.

**OCLC** *Online Computer Library Center*

**ODBC** *Open Database Connectivity* - Conectividad de base de datos abierta

**OHSAS** *Occupational Health and Safety Assessment Series*

**OLAP** *Online Analytical Processing* – Procesamiento analítico en línea

**OLTP** *Online Transaction Processing* – Procesamiento de transacciones en línea

**OPAC** *Online Public Access Catalog* - Catálogo Público de Acceso en Línea

**PCA** *Principal Component Analysis* - Análisis de componentes principales

**PFCE** Programa de Fortalecimiento de la Calidad Educativa

**PIFI** Programa Integral de Fortalecimiento Institucional

**PNPC** Programa Nacional de Posgrados de Calidad

**PRODEP** Programa para el Desarrollo Profesional Docente, para el Tipo Superior

**PROFOCIE** Programa de Fortalecimiento de la Calidad en Instituciones Educativas

**PubMed** Base de datos electrónica mantenida por la Biblioteca Nacional de Medicina de Estados Unidos

**RAE** Real Academia Española

**RDA** *Resource Description and Access* – Descripción y acceso a recursos

**RDBMS** *Relational Database Management System* - Sistema de administración o gestión de bases de datos relacionales

**RRSS** Redes sociales

**SCDD** Sistema de Clasificación Decimal Dewey

**SEGA** Sistema Estratégico de Gestión Académica de la UACH

**SEP** Secretaría de Educación Pública

**SGBD** Sistema de Gestión de Bases de Datos

**SGU** Sistema de Gestión Universitario (referente a la calidad en ciertos procesos de la UACH)

**SIGB** Sistema Integrado de Gestión Bibliotecaria. También conocido como Sistema de Gestión de Biblioteca (del inglés *Library Management System* o *Integrated Library System*)

**SQL** *Structured Query Language* – Lenguaje de consulta estructurada



**SUAE** Sistema Universitario de Administración Escolar

**SUBA** Sistema Universitario de Bibliotecas Académicas

**TDyE** Toma de decisiones y evaluación

**UACH** Universidad Autónoma de Chihuahua

**UI** *User Interface* - Interfaz de usuario

**UNAM** Universidad Nacional Autónoma de México

**UNIQ** Sistema de gestión de información documental del Sistema de Gestión de Calidad y Sistema de Gestión Universitario de la UACH

**UTF-8** *8-bit Unicode Transformation Format* - Formato o estándar de codificación de caracteres Unicode e ISO 10646 que utiliza símbolos de longitud variable.

**UX** *User eXperience* - Experiencia del usuario

**XML** *eXtensible Markup Language* - Lenguaje de Marcado Extensible

## Lista de tablas

Tabla 1. Comparación de sistemas OLTP y OLAP .....	16
Tabla 2. Tareas de minería de datos .....	19
Tabla 3. Listado de bibliotecas que conforman el SUBA .....	38
Tabla 4. Programas de estudio ofrecidos en las unidades académicas que atiende la BDESSalud.....	39
Tabla 5. Programas de estudio acreditados .....	40
Tabla 6. Cuerpos académicos de las unidades académicas atendidas por la BDESSalud.....	41
Tabla 7. Base de usuarios de la BDESSalud .....	43
Tabla 8. Normatividad bajo la cual opera el SUBA .....	47
Tabla 9. Metas e indicadores del Proceso General de Servicios Bibliotecarios del SUBA.....	48
Tabla 10. Principales temas y subtemas tratados en reuniones del CCSUBA.....	51
Tabla 11. Datos relativos a las Estadísticas Generales de la BDESSalud.....	55
Tabla 12. Indicadores y perspectivas de las preguntas de negocio de la BDESSalud.....	65
Tabla 13. Resumen de las fuentes de datos de la BDESSalud.....	70
Tabla 14. Mapeo de las fuentes de datos de las perspectivas del modelo lógico pertenecientes a la dimensión MDOC .....	77
Tabla 15. Mapeo de las fuentes de datos de la perspectiva del modelo lógico perteneciente a la subdimensión TEMA_MDOC .....	78
Tabla 16. Mapeo de las fuentes de datos de las perspectivas del modelo lógico pertenecientes a la dimensión USUARIO .....	79
Tabla 17. Mapeo de las fuentes de datos de la perspectiva del modelo lógico perteneciente a la subdimensión TIPO_USUARIO.....	80
Tabla 18. Mapeo de las fuentes de datos de la perspectiva del modelo lógico perteneciente a la subdimensión CARRERA.....	81
Tabla 19. Mapeo de las fuentes de datos de la perspectiva del modelo lógico perteneciente a la subdimensión UNIDAD_ACADEMICA .....	82
Tabla 20. Mapeo de las fuentes de datos de la perspectiva del modelo lógico perteneciente a la subdimensión BIBLIOTECA_ADSCRIPCION .....	83
Tabla 21. Mapeo de las fuentes de datos de la perspectiva del modelo lógico perteneciente a la dimensión TIPO_TRANSACCION.....	84
Tabla 22. Mapeo de las fuentes de datos de las perspectivas del modelo lógico pertenecientes a la dimensión TIPO_DEVOLUCION .....	85
Tabla 23. Mapeo de las fuentes de datos de las perspectivas del modelo lógico pertenecientes a la dimensión BIBLIOTECARIO .....	86
Tabla 24. Mapeo de las fuentes de datos de las perspectivas del modelo lógico pertenecientes a la dimensión MULTA .....	87
Tabla 25. Mapeo de las fuentes de datos de las perspectivas del modelo lógico pertenecientes a la dimensión TIEMPO .....	88
Tabla 26. Mapeo de las fuentes de datos de los indicadores del modelo lógico pertenecientes a la tabla de hechos CIRCULACION_MDOC.....	89

<b>Tabla 27. Tipos de datos, características y funcionalidades .....</b>	<b>109</b>
<b>Tabla 28. Clasificación – Desagregado de temas para el MDoc de la BDESSalud .....</b>	<b>112</b>
<b>Tabla 29. Extracto de la tabla para la asignación de unidad académica (facultad), nivel académico y programa educativo del usuario .....</b>	<b>113</b>
<b>Tabla 30. Variables inherentes y transaccionales para la matriz de entrenamiento.....</b>	<b>150</b>
<b>Tabla 31. Temas con la mayor concentración de consultas en la BDESSalud .....</b>	<b>151</b>
<b>Tabla 32. Frecuencia de usuarios por Facultad.....</b>	<b>154</b>
<b>Tabla 33. Filtros empleados para la consulta de los indicadores.....</b>	<b>163</b>
<b>Tabla 34. Títulos más consultados por tipo de usuario y tema.....</b>	<b>166</b>
<b>Tabla 35. Transacciones realizadas por bibliotecario.....</b>	<b>169</b>
<b>Tabla 36. Actualidad de la colección del MDoc por área temática [fragmento] .....</b>	<b>173</b>
<b>Tabla 37. Actualidad del MDoc de acuerdo con las áreas temáticas exploradas.....</b>	<b>175</b>
<b>Tabla 38. Tipo de adquisición del material solicitado en préstamo (2010-2019) .....</b>	<b>178</b>
<b>Tabla 39. Multas canceladas por bibliotecario, según el tipo de usuario y carrera.....</b>	<b>181</b>
<b>Tabla 40. Devoluciones vencidas por tipo de usuario, según el tema del MDoc .....</b>	<b>184</b>
<b>Tabla 41. Principales características de los cuatro grupos identificados .....</b>	<b>188</b>
<b>Tabla 42. Diferencias en temas consultados según el grupo de usuarios.....</b>	<b>190</b>

## **Lista de figuras**

<b>Figura 1. Pirámide del conocimiento y de la gestión del conocimiento revisada .....</b>	<b>10</b>
<b>Figura 2. Relación entre dato, información y conocimiento .....</b>	<b>12</b>
<b>Figura 3. Taxonomía de técnicas y métodos de minería de datos .....</b>	<b>22</b>
<b>Figura 4. Roles de los participantes en equipos de Ciencia de datos y requerimientos de habilidades.....</b>	<b>25</b>
<b>Figura 5. Organigrama de la Universidad Autónoma de Chihuahua.....</b>	<b>36</b>
<b>Figura 6. Organigrama del Sistema Universitario de Bibliotecas Académicas (SUBA).....</b>	<b>37</b>
<b>Figura 7. Variables e indicadores del proceso de circulación de material documental .....</b>	<b>68</b>
<b>Figura 8. Modelo conceptual del proceso de circulación de material documental.....</b>	<b>69</b>
<b>Figura 9. Modelo conceptual ampliado del DW de la BDESSalud .....</b>	<b>94</b>
<b>Figura 10. Dimensión MDoc .....</b>	<b>96</b>
<b>Figura 11. Subdimensión tema MDoc.....</b>	<b>97</b>
<b>Figura 12. Dimensión usuario.....</b>	<b>97</b>
<b>Figura 13. Subdimensión tipo usuario .....</b>	<b>98</b>
<b>Figura 14. Subdimensión carrera .....</b>	<b>98</b>
<b>Figura 15. Subdimensión unidad académica .....</b>	<b>99</b>
<b>Figura 16. Subdimensión biblioteca de adscripción .....</b>	<b>99</b>
<b>Figura 17. Dimensión tipo de transacción .....</b>	<b>99</b>
<b>Figura 18. Dimensión tipo de devolución.....</b>	<b>100</b>
<b>Figura 19. Dimensión bibliotecario .....</b>	<b>100</b>
<b>Figura 20. Dimensión multa .....</b>	<b>101</b>
<b>Figura 21. Dimensión tiempo .....</b>	<b>101</b>

<b>Figura 22. Tabla de hechos del proceso de circulación del MDoc</b> .....	102
<b>Figura 23. Uniones de las tablas dimensionales y subdimensionales con la tabla de hechos del proceso de circulación del MDoc de la BDESSalud</b> .....	104
<b>Figura 24. Proceso ETL principal planteado para la carga inicial del DW BDESSalud</b> ....	115
<b>Figura 25. Carga de la subdimensión TIPO_USUARIO</b> .....	117
<b>Figura 26. Carga de la subdimensión UNIDAD_ACADEMICA</b> .....	118
<b>Figura 27. Carga de la subdimensión BIBLIOTECA_ADSCRIPCION</b> .....	119
<b>Figura 28. Carga de la subdimensión CARRERA</b> .....	120
<b>Figura 29. Carga de la subdimensión TEMA_MDOC</b> .....	122
<b>Figura 30. Carga de la dimensión USUARIO</b> .....	123
<b>Figura 31. Carga de la dimensión MDOC</b> .....	125
<b>Figura 32. Carga de la dimensión TIPO_DEVOLUCION</b> .....	128
<b>Figura 33. Carga de la dimensión TIPO_TRANSACCION</b> .....	129
<b>Figura 34. Carga de la dimensión MULTA</b> .....	131
<b>Figura 35. Carga de la dimensión BIBLIOTECARIO</b> .....	132
<b>Figura 36. Carga de la dimensión TIEMPO</b> .....	134
<b>Figura 37. Carga de la tabla de hechos CIRCULACION_MDOC</b> .....	138
<b>Figura 38. Componentes del Cubo DW BDESSalud en MySQL Workbench</b> .....	141
<b>Figura 39. Prototipo de interfaz de consulta y visualización del DW de la BDESSalud</b> ...	143
<b>Figura 40. Sexo del usuario</b> .....	153
<b>Figura 41. Tipo de usuario</b> .....	153
<b>Figura 42. Carrera</b> .....	154
<b>Figura 43. Promedio mensual de préstamos de MDoc</b> .....	156
<b>Figura 44. Promedio de préstamos en meses con actividad</b> .....	156
<b>Figura 45. Antigüedad promedio del MDoc consultado</b> .....	157
<b>Figura 46. Mayor antigüedad del MDoc consultado</b> .....	157
<b>Figura 47. Menor antigüedad del MDoc consultado (en años)</b> .....	158
<b>Figura 48. Promedio de días de retraso por préstamo</b> .....	158
<b>Figura 49. Porcentaje de devoluciones tardías</b> .....	159
<b>Figura 50. Número de grupos temáticos consultados</b> .....	159
<b>Figura 51. Gráfica de codo para determinar del número óptimo de grupos</b> .....	161
<b>Figura 52. Coeficiente de silueta para 4 grupos de usuarios</b> .....	162
<b>Figura 53. Títulos más consultados agrupados por tema, por tipo de usuario y carrera</b>	167
<b>Figura 54. Transacciones realizadas por bibliotecario</b> .....	170
<b>Figura 55. Actualidad de la colección de MDoc por área temática</b> .....	174
<b>Figura 56. Tipo de adquisición del MDoc solicitado en préstamo</b> .....	179
<b>Figura 57. Multas canceladas por bibliotecario, según el tipo de usuario y carrera</b> .....	182
<b>Figura 58. Devoluciones vencidas por tipo de usuario, según tema del MDoc, 1ª capa</b> ..	185
<b>Figura 59. Devoluciones vencidas por tipo de usuario, según tema del MDoc, 2ª capa</b> ..	186
<b>Figura 60. Esquema del modelo de bibliominería</b> .....	191

## Resumen

El propósito de esta investigación fue determinar estructuras y conjuntos de datos requeridos para identificar patrones de uso que generen conocimiento significativo, los cuales contribuyen a los procesos de evaluación y toma de decisiones en bibliotecas. Se manejó un enfoque metodológico mixto, en el cual a partir de la selección y reconocimiento de la unidad de análisis y observación (una biblioteca académica del área de ciencias de la salud perteneciente a un sistema de bibliotecas de una universidad pública estatal) se diseñó un modelo de bibliominería fundamentado en elementos tanto del paradigma cualitativo (a partir de la revisión y análisis documental, así como la realización de entrevistas semiestructuradas con actores de interés) como del paradigma cuantitativo (análisis estadístico de datos de una encuesta aplicada a profesionales de la información, análisis mediante distribuciones estadísticas a las características inherentes de los usuarios bajo estudio, así como la aplicación del algoritmo de agrupación de la minería de datos y visualización de datos obtenidos de un data warehouse diseñado bajo la metodología Hefesto). Su finalidad fue el diseño y propuesta de un modelo de bibliominería que permite a los tomadores de decisiones en bibliotecas universitarias definir perfiles de usuarios considerando sus patrones de uso de material documental impreso, para lo cual se emplearon datos de circulación almacenados por la unidad de análisis y observación correspondientes a las transacciones de material documental impreso realizadas por sus usuarios durante el periodo del 01 de enero de 2010 al 31 de diciembre de 2019.

**Palabras clave:** Bibliominería, data warehouse, bibliotecas académicas, bibliotecas universitarias, Ciencia de datos, minería de datos, algoritmo K-Means, segmentación de usuarios, evaluación, toma de decisiones.

## Abstract

The purpose of this research is to determine structures and data sets required to identify patterns of use that generate meaningful knowledge, which contribute to evaluation and decision-making processes in libraries. A mixed methodological approach was used, in which, based on the selection and recognition of the unit of analysis and observation (an academic library in the health sciences area, which belongs to a public state university's library system), a bibliomining model was designed based on elements of both the qualitative paradigm (based on documentary review and analysis, as well as conducting semi-structured interviews with stakeholders) and the quantitative paradigm (statistical analysis of data from a survey applied to information professionals, analysis through statistical distributions to the inherent characteristics of the users under study, as well as the application of the data mining grouping algorithm and data visualization obtained from a data warehouse designed under the Hefesto methodology). Its purpose was the design and proposal of a bibliomining model that allows decision makers in university libraries to define user profiles considering their patterns of use of printed documentary material. This research used circulation data stored by the academic library under study, corresponding to the transactions of printed documentary material conducted by its users during the period from January 1, 2010, to December 31, 2019.

**Keywords:** Bibliomining, data warehouse, academic libraries, university libraries, Data science, data mining, K-Means algorithm, user segmentation, evaluation, decision making.

# Introducción

La toma de decisiones juega un papel crucial en todo tipo de organización o sistema. Un buen desempeño en esta área refleja su capacidad para identificar información útil, confiable, válida y actual que le permita reconocer cabalmente la dinámica sobre las características de su entorno interno y externo, así como actuar en consecuencia y obtener resultados óptimos en su desarrollo. Las vías a partir de las cuales las organizaciones acceden a esta información son heterogéneas. Empero, un proceso que es empleado de manera recurrente es el de la evaluación continua de sus actividades y áreas de interés. Tal ejercicio les permite recabar datos útiles para habilitar la apropiada toma de decisiones.

De acuerdo con Rodríguez Cruz (2014), el posibilitar la toma de decisiones a nivel gerencial requiere llevar a cabo una adecuada gestión de información que consideraría los siguientes elementos:

- Identificación de las necesidades informativas a nivel estratégico.
- Identificación y organización de la información relevante a este nivel.
- Flujos informativos constantes de información operativa, gerencial y estratégica.
- Los recursos informacionales (materiales, tecnológicos) que aseguran contar con una infraestructura que facilite el manejo y uso de información.
- Dominio de los procesos informacionales que se desarrollan para poder percibir, crear conocimiento y elegir la mejor alternativa de decisión.
- Sistemas informativos de apoyo a la toma de decisiones.
- Claridad de los roles informacionales que asumen los decisores o individuos que conforman grupos de decisión.
- Competencias informacionales asociadas al manejo y uso de la información.

Este tipo de actividades no difiere en el ámbito de las unidades y actividades bibliotecarias y de la información. Así, los procesos valorativos en estas unidades y actividades han recurrido a toda suerte de metodologías tanto cualitativas, cuantitativas como mixtas para la obtención de información significativa y útil orientada a la toma de decisiones. Como ejemplos más comunes se acude a la observación simple; la observación participante; el monitoreo de las colecciones a partir de la recolección de datos de ingreso, uso y descarte de materiales; la revisión de costos y los efectos de la oferta de servicios bibliotecarios, así como la puesta en marcha de sondeos (a partir de encuestas, entrevistas y grupos de enfoque). En todo esfuerzo se han valido del desarrollo de normas, criterios, indicadores y estándares que buscan favorecer la reunión de datos, su comparación y análisis. De manera particular, cada unidad de información puede determinar cuál es el esquema mejor adaptado a sus características, así como a los objetivos y necesidades de evaluación que se haya trazado. Una buena selección bien implementada, le permitirá mejorar las operaciones del día a día.

Ahora, el formidable desafío al que se enfrentan las organizaciones implica la recopilación, gestión y presentación de sus datos para que la administración pueda tomar decisiones informadas y oportunas (Nemati y Barko, 2003, p. 283). Durante las últimas décadas han surgido un conjunto de metodologías y procesos que han potencializado las capacidades del uso de los datos bibliográficos, estadísticos y textuales, entre otros, en la búsqueda de patrones significativos en el descubrimiento de conocimiento útil para la toma de decisiones. Entre estas nuevas metodologías se destaca el Descubrimiento de Conocimiento en Bases de Datos (en adelante KDD por las siglas en inglés de *Knowledge Discovery in Databases*), que integra como uno de sus principales procesos a la minería de datos, a partir del cual surge un proceso más específico y orientado a la identificación de patrones de información en las actividades bibliotecarias denominado bibliominería (*bibliomining*).

La bibliominería es un proceso emergente que resulta de la combinación de la minería de datos (en adelante MD), la bibliometría, la estadística y de las herramientas usadas para extraer patrones de comportamiento basados en el uso de los recursos de un sistema de biblioteca. El término fue usado por primera vez por Nicholson y Stanton



(Nicholson y Stanton, 2003) como resultado de una discusión de la aplicación de la MD a las bibliotecas. Aunque en la literatura especializada muchos trabajos hacen referencia en este contexto al término biblioteca, éstos no se refieren a una biblioteca tradicional, sino que están haciendo alusión a las bibliotecas implementadas en determinado *software* o programa de cómputo, como los sistemas de recursos que acumulan grandes volúmenes de información en las bibliotecas especializadas (Nicholson, 2003a, p. 146).

De la definición anterior, aportada por los propios creadores del término, se concibe a la bibliominería como un proceso en el cual se deben considerar las siguientes etapas:

- Determinación de áreas focales o problemas específicos que requieren de la toma de decisiones en la biblioteca.
- Identificación de las fuentes de datos apropiadas.
- Creación del repositorio o almacén de datos (en adelante DW por las siglas en inglés de *data warehouse*) confeccionado a partir de los datos seleccionados de las bases de datos del sistema que contiene una versión de los datos limpios y reformateados para su análisis.
- Selección de las herramientas apropiadas para el análisis de los datos.
- Descubrimiento de patrones y creación de reportes a partir de la MD.
- Análisis e implementación de los resultados.

Estas etapas de la bibliominería son consideradas por el propio Nicholson como un proceso cíclico por naturaleza, es decir el descubrimiento de patrones de comportamiento en los datos pueden suscitar nuevas interrogantes que echen a andar el proceso de nuevo (Nicholson, 2003b). Los avances obtenidos en este frente de investigación por los autores de referencia motivan el interés por indagar en la propuesta y diseño de un modelo de bibliominería orientado a identificar patrones de uso que generen conocimiento significativo para la toma de decisiones y la evaluación de actividades bibliotecarias.

Como otras organizaciones, en las últimas décadas los sistemas relativos a las actividades bibliotecarias han experimentado un notable incremento en el

almacenamiento y la cantidad de datos originados por las diversas transacciones que se establecen entre sus tareas, procesos, recursos y servicios, tales como: circulación del material documental (tanto impreso como electrónico), desarrollo de colecciones, formación de recursos humanos, servicios a usuarios, presupuestos, identificación de organizaciones afines, por solo mencionar los más representativos. Varios visionarios se han percatado de la riqueza que este cúmulo de información puede ofrecer para comprender y aprovechar mejor las dinámicas establecidas en estos tipos de sistemas y actividades. Sin embargo, en el ámbito nacional ha habido contados intentos por explotar estos datos y aprovechar su potencial en los procesos de toma de decisiones y evaluación de su comportamiento. El surgimiento de la bibliominería, frente de nueva creación en este tipo de actividades, ya empieza a dar sus primeras aportaciones es este sentido.

De acuerdo con Candás Romero (2006), “el estudio de los datos y la información almacenados en las bases de datos ofrece una visión perspectiva (qué se está haciendo y cómo se está haciendo) y prospectiva (cómo puede evolucionar [el sistema] en un futuro a corto-medio plazo) de la organización, de ahí su importancia y actualidad como frente de investigación en apoyo a la toma de decisiones [y en los procesos de evaluación] (párr. 5).

A pesar de que generalmente las bibliotecas cuentan con sistemas que les permiten la generación de estadísticas de sus actividades diarias para llevar un control de éstas, es habitual que los datos que arrojan no tengan la posibilidad de integrarse u optimizarse para alcanzar su máximo potencial analítico. Es por lo anterior que este proyecto tiene como propósito general mejorar el uso, análisis y aprovechamiento de datos producidos por actividades bibliotecarias a partir de la aplicación de los principios y métodos de la bibliominería, herramienta que, aun cuando ofrece variadas aplicaciones en los sistemas de información, no ha sido explorada y empleada lo suficiente en el ámbito profesional nacional.

Luego de conocer los antecedentes ya descritos, planteamos el problema de investigación que interesa a esta tesis a partir de la siguiente pregunta:

*¿Cuáles son las estructuras y conjuntos de datos que se requieren para la identificación de patrones de uso que generen conocimiento significativo y en qué medida estos patrones descubiertos contribuyen a los procesos de evaluación y toma de decisiones de las actividades bibliotecarias?*

De esta manera, el objetivo general a lograr con esta investigación es el determinar las estructuras de datos, así como las metodologías y herramientas que propicien la identificación de patrones de uso que generen conocimiento significativo con el propósito de establecer su relación con los procesos de evaluación y la toma de decisiones en actividades bibliotecarias.

Los objetivos específicos son:

- Identificar las relaciones existentes entre los datos estadísticos de series cronológicas derivados de los procesos y servicios bibliotecarios y los requerimientos para la evaluación y toma de decisiones de una biblioteca universitaria, con el propósito de seleccionar los datos que integrarán el repositorio de datos diseñado con estos fines.
- Diseñar y estructurar un repositorio de datos orientado a la identificación de patrones de uso que brinden conocimiento significativo sobre procesos, tareas y operaciones unitarias de los servicios bibliotecarios.
- Proponer un modelo de bibliominería con el fin de explicar las relaciones teóricas y metodológicas existentes entre la estructura de datos mensurables que se generan o pudieran generar en actividades bibliotecarias y su capacidad de aportar información útil en los procesos de evaluación y toma de decisiones de tales actividades.
- Identificar qué datos no se generan y/o almacenan en el sistema, que podrían ser útiles para hacer otros estudios y obtener mayor conocimiento.

Las hipótesis por comprobar a partir de esta investigación son

- No existe una relación completamente funcional entre: 1) las estadísticas generadas por las actividades bibliotecarias y, 2) las necesidades mostradas en los procesos de evaluación y toma de decisiones de estas actividades.
- La bibliominería tiene una relación directamente proporcional con la calidad de la gestión bibliotecaria, dado que contribuye la optimización de los datos primarios y eleva su potencial de utilidad en los procesos de evaluación y toma de decisiones de las actividades bibliotecarias.
- Considerando lo anterior, resultaría factible el diseño de un modelo de bibliominería que garantice las relaciones teóricas y metodológicas necesarias en la identificación de patrones de uso que generen conocimiento significativo orientado a la evaluación y a la toma de decisiones de actividades bibliotecarias.

La presente investigación manejó un enfoque metodológico mixto, en el cual a partir de la selección y reconocimiento de la unidad de análisis y observación (una biblioteca académica perteneciente a un sistema de bibliotecas de una universidad pública estatal) se diseñó un modelo de bibliominería fundamentado en elementos tanto del paradigma cualitativo (a partir de la revisión y análisis documental, así como la realización de entrevistas semiestructuradas con actores de interés) como del paradigma cuantitativo (análisis estadístico de datos de una encuesta aplicada a profesionales de la información, análisis mediante distribuciones estadísticas a las características inherentes de los usuarios bajo estudio, así como la aplicación del algoritmo de agrupación de la minería de datos y visualización de datos). El tipo de investigación realizada es aplicada y descriptiva. Su finalidad fue el diseño y propuesta de un modelo de bibliominería que permita a los tomadores de decisiones en bibliotecas universitarias delimitar perfiles de usuarios considerando los patrones de uso de material documental impreso que éstos han trazado en los datos resguardados en sus sistemas integrales de gestión bibliotecaria. Para poner en marcha el modelo propuesto se emplearon los datos almacenados por la unidad de análisis y observación correspondientes a las

transacciones de material documental impreso realizadas por sus usuarios durante el periodo del 01 de enero de 2010 al 31 de diciembre de 2019.

La tesis se ha estructurado en cinco capítulos. En el primero de ellos se delimita el marco disciplinar y teórico conceptual que sustenta el trabajo de investigación a partir de la explicación de los elementos que caracterizan a las bibliotecas universitarias, la necesidad de encontrar aproximaciones que permitan usar de manera pertinente y óptima los datos que éstas tienen disponibles en aras de mejorar sus procesos y servicios, así como rendir cuentas sobre su funcionamiento. Se exploran las características multi, inter y transdisciplinarias que presentan los Estudios Métricos de la Información y de qué manera éstos se relacionan con la Ciencia de datos en aras de ampliar sus paradigmas metodológicos y cognitivos para encontrar nuevas vías de análisis útiles para los profesionales de la información, además de señalar varias aplicaciones de la Ciencia de datos en el ámbito bibliotecológico, tanto en el ámbito internacional como nacional. El segundo capítulo se considera fundamental en este trabajo, pues en él se describe la metodología empleada partiendo de la descripción de la unidad de análisis y observación, seguido de la explicación detallada del desarrollo del almacén de datos propuesto (en adelante DW, por las siglas de *data warehouse*) y del diseño y aplicación del ejercicio de bibliominería. El tercer capítulo muestra el análisis de resultados, el cual se divide en la revisión de las consultas hechas al almacén de datos creado y en la descripción de los resultados obtenidos a partir de la aplicación de bibliominería. El cuarto capítulo expone la discusión de los resultados obtenidos y su contraste con la literatura científica consultada. El quinto capítulo enlista las conclusiones derivadas del estudio realizado, de interés para aquellos que necesitan optimizar el uso de los datos almacenados por sus sistemas de gestión bibliotecaria y que para ello desean involucrarse en tareas de bibliominería.

# CAPÍTULO 1. Evaluación de la biblioteca universitaria mediante aplicaciones de la Ciencia de datos

“Whether you know it or not;  
whether you accept it or not;  
whether you are ready for it or not,  
data science is coming.”  
- Zhu y Xiong, 2015.

En las últimas décadas hemos presenciado un crecimiento exponencial de la información que se genera día a día en todo tipo de actividades. Kallinikos (2006) sugiere que, más allá de considerarse como un resultado de incidentes azarosos, la expansión del universo informativo es una característica intrínseca del mundo contemporáneo, asociada a la existencia de mecanismos de almacenamiento y actualización cada vez más sofisticados, a la mayor disponibilidad de recursos en línea y a la capacidad combinatoria o de combinar<sup>1</sup> la información tecnológica. “Es precisamente la posibilidad de yuxtaponer, comparar y combinar información actualizada sobre patrones de vida individuales a partir de sistemas y bases de datos lo que abre un nuevo espacio de oportunidades, donde la realidad puede ser percibida de manera expandida, lo cual, a su vez, promueve nuevos cursos de acción” (Kallinikos, 2006, p. 99).

Lo anterior constituye una visión prometedora en cuanto a las posibilidades de explotación de tal cúmulo informativo. En el caso de las bibliotecas, el adecuado manejo de la información que refleja sus actividades y resultados se erige como una tarea fundamental, pues, aparte de proveer insumos necesarios para continuar su funcionamiento, es también materia prima para llevar a cabo procesos valorativos que son necesarios para hacer frente a las exigencias que su entorno le plantea.

---

<sup>1</sup> Del término en inglés *combinability*.

En un ambiente en constante cambio, caracterizado por presupuestos cada vez más precarios y la demanda de mayor innovación y dinamismo en los servicios que prestan, los bibliotecarios deben documentar y evidenciar el valor de las bibliotecas académicas, así como cuáles son sus aportes para el logro de la misión y visión de los objetivos institucionales.

Considerando lo expuesto y tomando en cuenta que uno de los temas que de manera constante aparece en el ámbito académico es la necesidad de ofrecer servicios educativos de calidad, se puede vislumbrar la necesidad de las bibliotecas universitarias de buscar los medios para hacer un uso pertinente de la información que tienen disponible para mejorar continuamente sus procesos y servicios, así como el rendir cuentas sobre su funcionamiento y el impacto que éste tiene en el logro académico de sus diversas comunidades de usuarios.

## **1.1 Bibliotecas universitarias**

Dado que la unidad de análisis y observación en esta investigación está constituida por una biblioteca universitaria<sup>2</sup>, es necesario enumerar las características que estos centros de información poseen. Así, en principio es posible establecer que una biblioteca universitaria está delimitada en cuanto a su pertenencia a una entidad académica superior que la alberga, con la cual presenta una relación simbiótica permanente. El éxito que una biblioteca universitaria puede llegar a tener está relacionado al nivel de comprensión que los individuos que la integran tengan del ambiente universitario que los rodea, para lo cual deben conocer en primer lugar cuáles son las funciones básicas de la universidad. Wilson y Tauber (1963) señalan que dichas funciones son: la conservación del conocimiento y de las ideas; la enseñanza; la investigación; la publicación; el servicio de extensión y la interpretación de los resultados de investigación. Todas estas actividades continúan vigentes, aunque el énfasis otorgado a tales

---

<sup>2</sup> Este trabajo considera a las bibliotecas universitarias como bibliotecas de tipo académico.

elementos difiere entre las instituciones académicas. Esto marca pautas para distinguir unas de otras y, por ende, también para hacer distinciones entre sus bibliotecas.

Toda biblioteca, incluyendo a la universitaria, debe llevar a cabo una serie de funciones que podemos considerar como genéricas: localizar, seleccionar, compilar, organizar, administrar y preservar recursos informativos con el fin de difundirlos y transmitirlos a los usuarios a partir de los servicios que ofrece, para así satisfacer sus necesidades de información. Ahora, la universitaria o académica se diferencia de las bibliotecas pública, escolar, especializada y otras a partir de dos elementos clave:

- **Su forma de gobierno:** Cada biblioteca pertenece, invariablemente, a una organización de mayor jerarquía, por lo cual queda bajo la autoridad de instancias que operan fuera de ella. En el caso de las bibliotecas universitarias, las organizaciones macro de las que se desprenden pueden ser universidades, tecnológicos o escuelas de educación superior; y,
- **La comunidad a la que sirve:** Cada biblioteca existe para proveer servicios de información a una comunidad específica, lo cual significa que sus servicios, colecciones y accesos son desarrollados tomando en cuenta a dicha comunidad. Para las bibliotecas universitarias, ésta se conformaría por los docentes, estudiantes, investigadores y personal administrativo de la institución de que forma parte (Budd, 1998, p. 3).

De acuerdo con la American Library Association (ALA), la biblioteca universitaria es una “biblioteca (o sistema de éstas) establecida, mantenida y administrada por una universidad para cubrir las necesidades de información de sus estudiantes y apoyar sus programas educativos, de investigación y demás servicios” (Young, 1988). Por su parte, Garza Mercado (2003, p. 179) indica que “la biblioteca de enseñanza superior tiene por objeto apoyar las funciones de docencia, investigación, difusión y vinculación (o a los requerimientos nacionales, regionales y estatales) de la universidad, el instituto tecnológico o la escuela normal de la que forma parte”. En los últimos años las funciones que realizan las bibliotecas universitarias se han adaptado a las demandas de sus usuarios, por lo cual les ha sido necesario llevar a cabo actividades de tipo educativo a



partir de las cuales buscan desarrollar las habilidades y competencias de éstos en el uso y manejo de la información, así como el transmitirles conocimientos referidos a las metodologías de la investigación y al análisis y visualización de datos.

El Consejo Nacional para Asuntos Bibliotecarios de las Instituciones de Educación Superior (CONPAB-IES) indica en su normatividad que, “[p]ara armonizar los elementos que intervienen en el proceso de organización y operación de la biblioteca académica, se plantea que ésta funcione como un sistema; con una visión estratégica, que ofrezca servicios y recursos de información en los que prevalezca la calidad y la evaluación; con la inclusión de estándares e indicadores, donde la comunicación y el cambio sean premisas fundamentales de la cultura organizacional” (CONPAB-IES 2012, pp. 14-15).

La biblioteca universitaria deberá contar con un documento institucional o estatutario, un reglamento o algún otro elemento que le sirva al bibliotecario como fuente de autoridad para ratificar las funciones bibliotecarias y que defina legalmente los diferentes recursos de la biblioteca, así como para señalar la línea de autoridad que la rige, así como otras cuestiones de interés (Litton, 1974, pp. 35-36).

En cuanto a su organización interna, las bibliotecas universitarias pueden recurrir a principios administrativos como la departamentalización o segmentación por áreas especializadas, lo cual estará determinado según los recursos y características de cada biblioteca. En este rubro, Garza Mercado (2003, p. 27) señala que “En la organización de la biblioteca se emplean los mismos criterios que determinan la estructura de otras organizaciones: función, producto, clientela o derechohabiente y área geográfica”. Con lo anterior, puede haber departamentalización:

- Según el **tipo de procedimientos**: procesos técnicos (desarrollo de colecciones, catalogación y clasificación, etc.) y servicios públicos (consulta, circulación, fotocopiado, etc.).
- De acuerdo con **contenido de los materiales**: por ejemplo, por disciplinas (humanidades, ciencias sociales, ciencias puras, tecnología o ramas específicas).
- Según el **formato de los materiales** (audiovisuales, fuentes electrónicas, microformas).

- Tomando en cuenta los **requerimientos o ubicación de los usuarios** (Garza Mercado, 2003, pp. 27-28).

Sin importar la organización administrativa que implemente la biblioteca, están predefinidas las actividades básicas que ésta debe llevar a cabo, como lo son las actividades administrativas primarias, el desarrollo de colecciones, el proceso técnico de material documental, el préstamo de materiales, la orientación a usuarios, los servicios de referencia y la formación de usuarios, entre otras.

Como se puede observar con lo expuesto hasta el momento, las bibliotecas universitarias constituyen sistemas con un nivel de complejidad tal, que requieren de mecanismos apropiados que permitan valorar la manera en la cual se desempeñan. Para ello, han existido varias aproximaciones en cuanto a procesos de evaluación, mismos que han considerado diversas metodologías: cuantitativas, cualitativas y mixtas (Véase apartado 1.3 Antecedentes metodológicos de la evaluación de bibliotecas universitarias, en Mendoza Guillén, 2007, pp. 20-28). En el caso de esta tesis, se hará énfasis en los referentes valorativos de bibliotecas universitarias que están relacionados al tema de la minería de datos. Empero, para llegar a ello, primero se esbozará un panorama de cómo la Ciencia de datos ha llegado a permear dentro del área bibliotecológica, particularmente en el tema de la evaluación bibliotecaria.

## **1.2 Estudios Métricos de la Información**

Los Estudios Métricos de la Información (en adelante EMI) constituyen un campo multi, inter y transdisciplinario en el cual interactúan diversos métodos y modelos matemáticos y estadísticos con las disciplinas que integran el denominado Sistema de Conocimientos Bibliológico Informativo (Bibliotecología, Bibliografología, Archivología y Ciencia de la Información). A esta interacción se le atribuye el surgimiento de un conjunto de especialidades métricas (Bibliotecometría, Bibliometría, Archivometría e Informetría)<sup>3</sup>,

---

<sup>3</sup> “así como otras relacionadas con ellas como la cienciometría, la webmetría y más recientemente la altmetría” (Gorbea Portal, 2016, p. 11).

las cuales han alcanzado cierta notoriedad, unas más que otras, en la literatura especializada en los últimos veinte años (Gorbea, 2013, p. 16).

En cuanto al objetivo de este trabajo, resulta de particular interés la especialidad métrica denominada bibliotecometría. Ésta se considera como la equivalencia del término en inglés *librametry*, empleado por S. R. Ranganathan en 1948 para designar la aplicación de técnicas matemáticas y estadísticas a los problemas bibliotecarios (Sengupta, 1992). La bibliotecometría es definida por Setián y Gorbea (1990) como “la aplicación de métodos y modelos matemáticos al estudio de los fenómenos propios de la actividad bibliotecaria, con el fin de caracterizar el comportamiento de los componentes que integra[n] esa actividad y las tendencias que se presentan en el uso de las bibliotecas y sus fondos” (Gorbea Portal, 1994, p. 24).

A pesar de que la bibliotecometría no es una especialidad métrica a la cual se haga referencia de manera frecuente, de acuerdo con Hood y Wilson (2001) “puede ser útil conservar los términos ‘librametrics’ o ‘librametry’ [bibliotecometría] para estudios que no analizan específicamente la literatura, o al menos no están específicamente enfocados en los objetivos de la bibliometría y de la recuperación de información. [Tales estudios] incluyen análisis de la circulación de libros [...], de la superposición de la colección en bibliotecas [...], de adquisiciones [...], de política de multas [...] y de la distribución de la estantería [...], para lo cual emplean frecuentemente técnicas de optimización de la investigación de operaciones”.

A decir de Gorbea Portal (1998), “[l]a matematización de los fenómenos sociales estudiados por la bibliotecología [...] no deberá convertirse en el fin primordial [de las actividades bibliotecarias y de información]. [S]in embargo, el bibliotecólogo o especialista en información que pueda contar con herramientas matemáticas y estadísticas en el estudio o desempeño de sus funciones [...] tendrá ante sí un mundo infinito y casi inexplorado, que de llegar a dominarlo no sólo podrá contar con una información más objetiva para la toma de decisiones, sino que con ello estará contribuyendo también al desarrollo del cuerpo teórico de la disciplina” (p. 21).

Ahora bien, la metría de la información sigue siendo considerada “como campo de investigación emergente y de frontera de las Ciencias Bibliotecológica y de la Información” y “ha tenido un desarrollo acelerado en los últimos tiempos [debido a] su naturaleza multi, inter y transdisciplinaria” (Gorbea, 2013, p. 13). Lo anterior es claramente visible al percatarnos que hoy día las capacidades de esta área de conocimiento se ven potenciadas por aplicaciones derivadas de la Ciencia de datos y las ciencias de cómputo, todas ellas vinculadas al tratamiento, almacenamiento, procesamiento, análisis y visualización de datos. Con esto podemos detectar una muy interesante y desafiante área de oportunidad en la cual, a partir del empleo de nuevos enfoques, es factible para los profesionales de la información el coadyuvar a la consolidación de la bibliotecometría como especialidad métrica en lo particular y a la bibliotecología en lo general. A continuación, se hace una semblanza de aquellas áreas del conocimiento con las cuales los EMI han convergido con el propósito de ampliar sus paradigmas y apoyar al desarrollo de la ciencia bibliotecológica y, particularmente del interés del presente estudio, en el surgimiento de nuevas metodologías fundamentadas en el análisis de datos para la toma de decisiones y la evaluación de actividades bibliotecarias.

### **1.3 Ciencia de datos**

Desde los albores del siglo 21 ha sido evidente la explosión de información derivada del registro de numerosas y diversas actividades en el mundo digital. Los avances tecnológicos han permitido que el procesamiento y almacenamiento de datos sea cada vez más económico y eficiente, lo cual favorece su ingente acumulación. Los datos, alguna vez definidos como la representación numérica de alguna medida, se han convertido hoy en día en una unidad atómica que se puede capturar, medir, ver o escuchar, y, por tanto, [son susceptibles de ser] extraídos, analizados y convertidos en información y, finalmente, en nuevo conocimiento (Lewis y McGrath, 2019, p. 1). Todo esto es posible gracias a la Ciencia de datos (del inglés, *data science*). El uso del término *data science* o Ciencia de datos se remonta a 1960, con las aportaciones de John W.

Tukey, quien publicó el libro *The future of data análisis* (Ahmad, Hamid y Ahmed, 2022, p. 95), y el científico informático Peter Naur<sup>4</sup>. Sin embargo, la Ciencia de datos también tiene raíces evolutivas en la estadística. Desde la década de los setenta, los métodos matemáticos, estadísticos y de la ciencia de la computación o informática han contribuido a su desarrollo. Autores como Usama M. Fayyad, Gregory Piatetsky-Shapiro y Padhraic Smyth, entre otros, han coadyuvado enormemente a darle difusión mediante la organización de talleres, conferencias y la publicación de libros y artículos de importancia central para esta área del conocimiento (Lewis y McGrath, 2019, pp. 1-2).

De acuerdo con Zhu y Xiong (2015), la Ciencia de datos es una nueva ciencia que proporciona un innovador método de investigación a las ciencias naturales y sociales a partir del uso intensivo de datos. A decir de los autores, se espera que la Ciencia de datos ayude a mejorar los métodos actuales de investigación científica, a fin de formar nuevos métodos y desarrollar teorías, métodos y tecnologías específicas [y útiles] en diversos campos [del saber] (p. 4).

La Comunidad Global de Gestión de Datos (DAMA) señala en su *Data Management Body of Knowledge* (DAMA International, 2017) que la Ciencia de datos ha sido considerada como un término de moda desde principios de la década del 2000. El concepto y sus implicaciones “no son comprendidos, o al menos existe poco consenso con respecto a su significado”. La Ciencia de datos ha existido por largo tiempo: “solía llamársele estadística aplicada”. Empero, hasta hace poco, “el análisis a profundidad de los datos había sido limitado por la tecnología”. La capacidad para explorar patrones de datos ha evolucionado rápidamente en el último siglo con el arribo del *Big Data*<sup>5</sup> y las tecnologías que lo apoyan (DAMA International, 2017, p. 497). La Ciencia de datos combina la extracción, la minería de datos, el análisis estadístico y el aprendizaje automático con capacidades de integración y modelado de datos, para construir modelos

---

<sup>4</sup> Peter Naur abogaba por el uso del término datología (*datalogy*) o ciencia de los datos, a la cual definía como la ciencia de la naturaleza y uso de los datos. Inclusive propuso al nuevo vocablo como reemplazo al de informática. (Naur, 1966, p. 485).

<sup>5</sup> Término utilizado para describir extensos conjuntos de datos con características estructurales grandes y complejas. De acuerdo con los principales expertos en *Big Data*, un conjunto de datos debe cumplir con las 5V para ser considerado parte de esta área de estudio: volumen, velocidad, veracidad, valor y variedad (Al-Barashdi y Al-Karousi, 2018, p. 3).

predictivos que exploren patrones de contenido. Y para el desarrollo y valoración de tales modelos predictivos, el analista emplea el método científico (DAMA International, 2017, p. 498). En sí, el objetivo principal de la Ciencia de datos es la aplicación de diversos métodos y técnicas para la obtención de conocimiento a partir del análisis de datos.

Como una ciencia creciente y en evolución, la Ciencia de datos es un amplio campo de estudio con múltiples vertientes, desde el estudio de *Big Data*, inteligencia artificial, aprendizaje de máquinas, hasta el KDD y la minería de datos (los dos últimos son elementos de interés para la realización de esta tesis). De acuerdo con Burton y Lyon (2017, p. 33), la Ciencia de datos ha transformado los sectores del comercio, la salud y el gobierno y continuará influyendo en la transformación de otros sectores, sobre todo aquellos que busquen optimizar sus procesos y mejorar su toma de decisiones. El sector educativo, del cual dependen las bibliotecas académicas, no es la excepción.

#### 1.4.1 Descubrimiento de conocimiento a partir de bases de datos

Antes de entrar en materia sobre el descubrimiento de conocimiento a partir de bases de datos o KDD, es importante hacer algunas observaciones sobre los términos involucrados. En específico, se hará referencia al proceso de transformación del dato a la información, de la información al conocimiento y del conocimiento a la sabiduría (o DIKW, por sus siglas en inglés de *data, information, knowledge* y *wisdom*). Éste ha sido un tema estudiado de antaño, sobre todo en el área de la gestión del conocimiento. Aquí se emplearán los términos propuestos por Jennex y Bartczak (2013, p. 27):

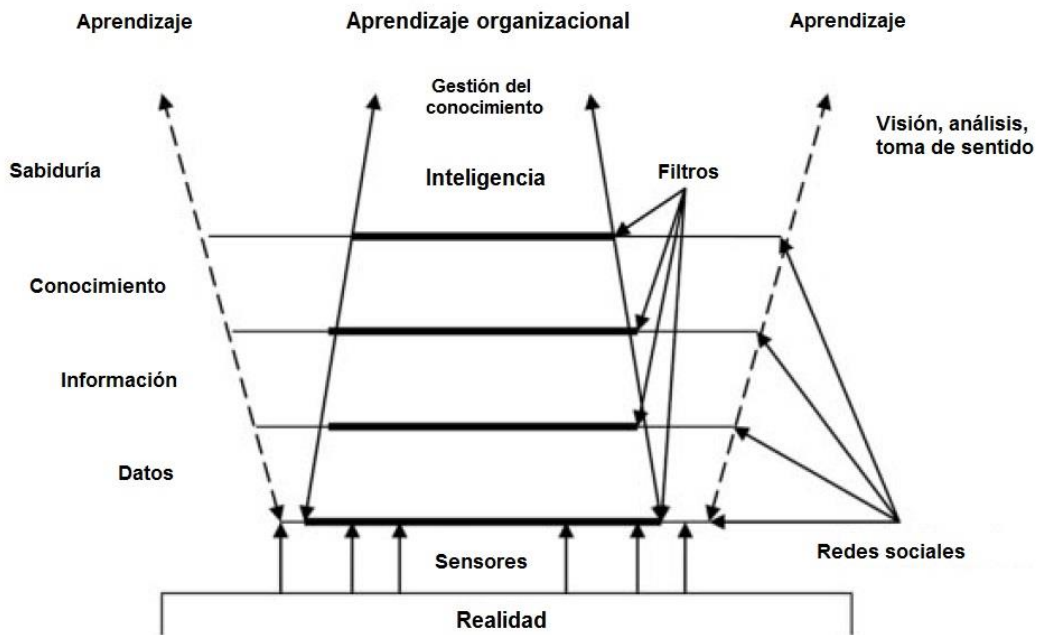
**Dato:** Hechos básicos, discretos, objetivos, como quién, qué, cuándo y dónde, sobre algo.

**Información:** Datos que se relacionan entre sí a partir de un contexto tal que proporcionan una historia o noticia útil.

**Conocimiento:** Información que ha sido entendida culturalmente de tal manera que explique el cómo y por qué sobre algo, o que proporciona una comprensión sobre algo.

**Sabiduría:** Es el colocar el conocimiento en un marco que permita su aplicación a situaciones diferentes y no necesariamente intuitivas.

Jennex y Bartczak (2013) aportan una nueva descripción gráfica de esta transformación, que se muestra en la Figura 1.



**Figura 1. Pirámide del conocimiento y de la gestión del conocimiento revisada**

Fuente: Jennex y Bartczak (2013, p. 25). (Traducción propia).

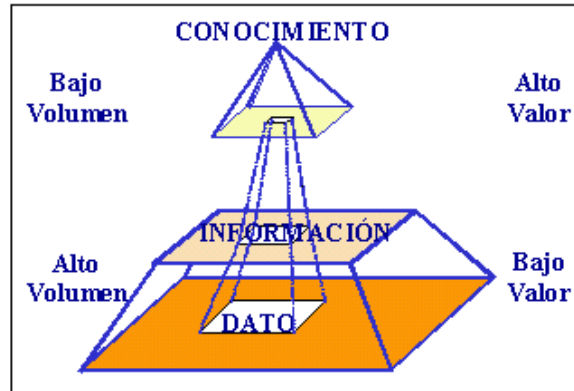
En esta propuesta, se indica la existencia de sensores que detectan, interpretan y capturan datos sobre la realidad circundante (los sensores pueden ser desde los sentidos humanos hasta elementos mecánicos o electrónicos). Las líneas punteadas reflejan el procesamiento de los datos en información, conocimiento y sabiduría empleando los procesos de visión, análisis y toma de sentido. Las líneas son bidireccionales, lo cual indica que la generación de información, conocimiento y sabiduría puede avanzar hacia la parte superior de la jerarquía, o bien, brindar retroalimentación hacia la parte inferior de la misma, a medida que el usuario comprende más de lo que está aprendiendo. Las redes sociales (RRSS) se emplean para hacer referencia a cualquier método (formal o informal, directo o indirecto) empleado para transferir datos, información, conocimiento y

sabiduría entre usuarios. El final superior de las flechas lleva al aprendizaje, considerado por los creadores del modelo como la adquisición de *DIKW* que lleva a un cambio de comportamiento o expectativas dentro del individuo o grupo que está inmerso en el aprendizaje (Jennex y Bartczak, 2013, p. 22).

Este planteamiento hace evidente que el apropiado uso de los datos es necesario para activar todo un proceso que lleva al aprendizaje o a una mayor comprensión sobre un fenómeno dado. En el caso de las organizaciones, se espera que tal aprendizaje conduzca a una mejora sustantiva de sus actividades y resultados, gracias a que existe un mayor conocimiento disponible para la toma de decisiones que le procuran ventajas competitivas.

Lo expuesto hasta el momento constituye el objetivo central del KDD: el descubrimiento de conocimiento [que involucra] el análisis de nuevos patrones [de datos] encontrados (o información) y la aplicación a éstos bajo cierto contexto, cultura y otros conocimientos previos, para luego pasar a la utilización de este conocimiento en la tarea de resolver problemas existentes. Otra manera de explicar esta transformación es la propuesta por Molina (1998), quien señala que, de forma general, en el proceso de extracción de conocimiento de bases de datos, los datos constituyen la materia prima bruta. En el momento en que el usuario atribuye algún significado especial a los datos, éstos pasan a ser información. Cuando los especialistas del dominio del problema elaboran una norma o regla, la interpretación de confrontación entre ese hecho (información) y esa regla constituye un conocimiento. En la Figura 2 podemos observar cuál es la jerarquía establecida entre los datos, la información y el conocimiento existente en una base de datos, de acuerdo con el volumen de estos elementos y al valor que los responsables de la toma de decisiones le otorgan a cada uno de ellos.





**Figura 2. Relación entre dato, información y conocimiento**

Fuente: Molina (1998, p.6).

Pasemos ahora a delimitar el descubrimiento de conocimiento a partir de bases de datos o KDD, el cual es el “proceso<sup>6</sup> no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia, comprensibles a partir de los datos” (Fayyad, Piatetsky-Shapiro y Smyth, 1996. Traducción propia). Esta tesis tiene como objetivo principal el obtener patrones de conocimiento valioso que permita coadyuvar a la toma de decisiones en una biblioteca universitaria. Pero ¿qué es un patrón de conocimiento? De manera general, un patrón puede entenderse como un modelo que sirve de muestra para sacar otra cosa igual (RAE, 2018). Empero, desde el punto de vista del KDD, un patrón puede ser interpretado como la ejemplificación de un modelo. La minería de datos (uno de los principales pasos de la metodología de KDD y a menudo tomado como sinónimo de ésta<sup>7</sup>) implica ajustar modelos a, o determinar patrones a partir de, datos observados<sup>8</sup>. De acuerdo con Raina y Krishnamurthy (2022, p. 33. Traducción propia), un modelo tiene dos propósitos: explicar<sup>9</sup> las observaciones de un fenómeno o hecho

<sup>6</sup> Fayyad, Piatetsky-Shapiro y Smyth vislumbraban al KDD como un campo emergente [del conocimiento] y como un proceso, empero, en el marco de esta tesis se considera al KDD como una metodología que engloba varios procesos (técnicas y herramientas), los cuales le permiten lograr la identificación de patrones en los datos.

<sup>7</sup> Como ocurre con Han, Kamber y Pei (2012, p. 1).

<sup>8</sup> Para la Encyclopedia Britannica, el reconocimiento de patrones “puede implicar la medición del objeto [o bien, el reconocimiento de los datos] para identificar atributos distintivos, la extracción de características para los atributos definitorios y la comparación con patrones conocidos para determinar una coincidencia o una falta de coincidencia” (Britannica, 2021).

<sup>9</sup> Descifrar las verdades o causas subyacentes del evento o fenómeno bajo estudio.

dado y [pronosticar]<sup>10</sup> futuras observaciones de éste. Los modelos ajustados desempeñan el papel del conocimiento inferido: si los modelos reflejan o no un conocimiento *útil* o *interesante*, ya es parte del proceso global e interactivo de KDD el determinarlo<sup>11</sup>, y en éste generalmente se requiere la participación del juicio humano (subjetivo). Hay dos formalismos matemáticos empleados en las tareas de ajuste de modelos: el enfoque estadístico/probabilístico, que permite efectos no deterministas en el modelo; y el modelo lógico, que es puramente determinista y no admite la posibilidad de incertidumbre en el proceso de modelado. El enfoque más empleado en las aplicaciones prácticas de minería de datos es el estadístico/probabilístico, debido a la incertidumbre típica sobre la naturaleza exacta de los procesos de generación de datos en el mundo real (Fayyad, Piatetsky-Shapiro y Smyth, 1996, pp. 11-12).

De acuerdo con Han, Kam y Pei (2012, p. 6), [la metodología] KDD es una secuencia iterativa de los siguientes pasos<sup>12</sup>:

1. **Limpieza de datos** (remoción de ruido y datos inconsistentes)
2. **Integración de datos** (combinación de múltiples fuentes de datos)
3. **Selección de datos** (se extrae de la(s) base(s) de datos fuente(s) los datos relevantes para el análisis)
4. **Transformación de datos** (los datos se consolidan en formas apropiadas realizando operaciones de resumen y agregación que permiten su minado)
5. **Minería de datos** (proceso esencial donde se aplican métodos inteligentes para extraer patrones de datos)

---

<sup>10</sup> Precisar la(s) futura(s) ocurrencia(s) del evento o fenómeno bajo estudio. El término empleado por el autor es *predict*, que puede traducirse como predecir. Empero, se prefirió emplear el término pronosticar (del inglés *to forecast*) y evadir predecir, que de acuerdo con la Real Academia Española significa “anunciar por revelación, conocimiento fundado, intuición o conjetura de algo que ha de suceder”, pues el término tiende a implicar más la adivinanza por revelación o intuición divina. En tanto, y a pesar de que pronosticar se utiliza como sinónimo de predecir, en ciencia se pronostica a partir de indicios o evidencias científicas extraídas del conocimiento y/o del análisis de los datos.

<sup>11</sup> En el paso final del KDD se enfatiza que el valor del conocimiento descubierto reside en su apropiado uso. [El conocimiento] debe consolidarse y resolverse con conocimientos previos, compartirse, reportarse, difundirse y aplicarse. Empero, para ello se requiere la integración adecuada del conocimiento descubierto, el conocimiento del dominio (o conocimiento del campo o tema bajo estudio) e intención estratégica para llegar a soluciones que no eran evidentes en los datos en sí (Uthurusamy, 1996, pp. 567-568).

<sup>12</sup> A los cuales podríamos anteponer la etapa de identificación y recopilación de datos.

6. **Evaluación de patrones** (identificación de patrones realmente interesantes que representan el conocimiento, a partir de medidas de interés)
7. **Presentación del conocimiento obtenido** (técnicas de visualización y representación del conocimiento son empleados para dar a conocer los resultados obtenidos a los usuarios finales) (Han, Kamber y Pei, 2012, pp. 7-8).

Dentro de la industria de la información, existe la tendencia de llevar a cabo la limpieza e integración de datos como una etapa de preprocesamiento, en la cual los datos resultantes se almacenan en un DW o almacén de datos. Más adelante se describirán las características de tal repositorio, así como su importancia en la labor del análisis intensivo de datos.

#### 1.4.2 Data warehouse

En la metodología de KDD, el concepto de data warehouse (en adelante DW) toma vital importancia, al ser la herramienta clave en la que los datos son sometidos a un tratamiento previo que posibilita su posterior análisis y aprovechamiento. El propósito principal de los almacenes de datos (DW) es permitir la ejecución rápida de consultas analíticas complejas usando lenguaje de consulta estructurada (en adelante *SQL*, por las siglas en inglés de *Structured Query Language*) (Raina y Krishnamurthy, 2022, p. 168). Su diseño y puesta en marcha involucra tareas como la extracción, transformación, limpieza, integración y carga de datos que son desheredados de bases de datos operacionales de las cuales proceden.

El término fue acuñado en 1988 por Barry A. Devlin y Paul T. Murphy en su artículo “*An architecture for a business and information system*”, en el cual analizan una nueva arquitectura de software que describen como un “*business data warehouse*” o almacén de datos de negocios. Empero, el crédito en cuanto a la difusión del término le pertenece al científico de cómputo William Harvey Inmon, quien ya en la década de 1970 comenzó a discutir los principios que subyacen a la creación de estos repositorios y en 1992 publicó el primer libro sobre el tema: *Building the data warehouse* (Tantawi, 2013). Inmon

define el DW como una recopilación de datos orientada a temas [de interés], integrada, con perspectiva histórica-temporal y no volátil, la cual respalda el proceso de toma de decisiones de una administración (Han et al. 2012, p. 126). Por otro lado, otro importante precursor de los DW, Ralph Kimball, define al término como un repositorio de información integrada proveniente de fuentes distribuidas, autónomas y posiblemente heterogéneas, específicamente estructurado para llevar a cabo análisis y consultas definidas.

Precisamente, las dos metodologías más importantes para el diseño y puesta en marcha de DW los han formulado Inmon y Kimball. Inmon, con un enfoque que va de arriba hacia abajo, indica que los datos deben extraerse de los sistemas de bases de datos transaccionales e integrarse a un DW centralizado (Inmon, 2003). En cambio, Kimball plantea un enfoque de abajo hacia arriba, en el cual varios *data marts*<sup>13</sup> (que representan cada uno a un proceso específico de negocio o a una dimensión) se unen para conformar el DW (Kimball y Ross, 2013).

Otra metodología alterna para el diseño e implementación de DW es la conocida como Hefesto, cuya primera versión fue dada a conocer por su autor, el ingeniero en sistemas argentino Dario Bernabeu, en 2007. Desde ese año a la fecha, Hefesto ha llegado a su versión número tres. En sí, esta metodología se distingue de otras por el hecho de que permite la construcción de DW de forma sencilla, ordenada e intuitiva, a partir del análisis de requerimientos de los usuarios, el análisis de las fuentes de datos a considerar, la creación del modelo lógico del DW y la integración de los datos que lo conformarán (Bernabeu y García, 2017).

Así, un DW es “un almacén de datos que recopila información de fuentes distintas, almacenándolas en una ubicación central para facilitar su acceso y análisis. Los datos almacenados constituyen un registro estático o una instantánea de cómo se veía cada elemento en un punto específico en el tiempo. Los datos no se actualizan [...] si la información ha cambiado con el tiempo, las instantáneas posteriores del mismo artículo simplemente se agregan al almacén de datos” (Tantawi, 2013). Puede decirse entonces

---

<sup>13</sup> Precisemos que un *data mart* es un subconjunto de un DW creado bajo los mismos principios, pero con un alcance más limitado.

que la información que contiene un DW es de tipo histórico. Su actualización se realiza a partir de cargas periódicas de datos provenientes de la fuente transaccional.

Un DW puede ser definido como un almacén o repositorio de datos que permite, valga la redundancia, el almacenamiento a largo plazo de datos recopilados de múltiples fuentes, almacenados bajo un esquema unificado, típicamente resumido, con perspectiva histórica y que por lo general reside en un solo sitio. Los datos son organizados de manera que se facilita la toma de decisiones, lo cual implica un proceso de limpieza, integración, transformación, carga y actualización periódica de datos. Estos sistemas de almacenamiento permiten llevar a cabo análisis de datos multidimensionales, proceso conocido también como procesamiento analítico en línea u OLAP (por sus siglas del inglés *Online Analytical Processing*). Por ello, los DW pueden considerarse como sistemas OLAP. Dado que existe una mayor familiaridad con los sistemas de bases de datos relacionales comerciales, resulta fácil comprender qué es un DW al compararlo con éstas. La principal tarea de un sistema de base de datos operacional en línea es llevar a cabo transacciones y procesamiento de consultas en línea. Estos sistemas se denominan sistemas de procesamiento de transacciones en línea u OLTP (por las siglas en inglés de *Online Transaction Processing*). Las operaciones en un DW (aplicación tipo OLAP) no son transacciones predominantes, como en las bases de datos operativas (aplicaciones OLTP), sino consultas complejas de unión, filtrado, agrupación y agregación de grandes cantidades de datos (Siguenza-Guzman, 2015, p. 9). En la Tabla 1 se muestra una comparación entre las principales características de los sistemas OLTP y OLAP.

**Tabla 1. Comparación de sistemas OLTP y OLAP**

<b>Característica</b>	<b>OLTP</b>	<b>OLAP</b>
<b>Propósito</b>	Procesamiento operativo (operaciones diarias)	Procesamiento de información (Recuperación de información, informes, análisis y minería de datos)
<b>Orientación</b>	Transaccional (orientado al cliente)	Análisis (orientado al mercado)

<b>Característica</b>	<b>OLTP</b>	<b>OLAP</b>
<b>Usuario</b>	Empleados operativos, administradores de bases de datos, profesionales de bases de datos, clientes	Trabajadores del conocimiento (Gerentes, administradores, ejecutivos, analistas)
<b>Función</b>	Operaciones diarias	Requerimientos informativos a largo plazo para la toma de decisiones
<b>Diseño de base de datos</b>	Modelo de datos entidad-relación. Diseño de bases de datos: orientado a la aplicación ( <i>Datos normalizados</i> )	Estrella/Copo de nieve, orientado a temas ( <i>Datos parcialmente desnormalizados, multidimensionales</i> ). Diseño de bases de datos: orientado a temas.
<b>Datos</b>	Actuales, actualización garantizada ( <i>Datos de funcionamiento de la organización, cambiantes, internos, incompletos</i> )	Históricos, precisión que se mantiene a lo largo del tiempo ( <i>Datos descriptivos, útiles para el análisis, el resumen, etc. Pueden ser internos y externos</i> )
<b>Resumen</b>	Primitivo, altamente detallado	Resumido, consolidado
<b>Visualización</b>	Detallado, plano relacional	Resumido, multidimensional
<b>Unidad de trabajo</b>	Transacción corta y simple	Consulta compleja
<b>Acceso</b>	Permite lectura/escritura ( <i>SQL</i> )	Permite lectura ( <i>SQL y herramientas propias, slice &amp; dice, drill, roll, pivot, etc.</i> )
<b>Enfoque</b>	Entrada de datos	Salida de información
<b>Operaciones</b>	Indización en llave primaria	Muchas exploraciones
<b>Número de registros accedidos</b>	Decenas	Millones
<b>Número de usuarios</b>	Miles	Cientos
<b>Tamaño de base de datos</b>	GB a GB de alto orden	≥ TB
<b>Prioridad</b>	Alto rendimiento, alta disponibilidad	Alta flexibilidad, autonomía del usuario final
<b>Métrica</b>	Rendimiento de transacción	Rendimiento de la consulta, tiempo de respuesta

Fuentes: Adaptación de Han, Kamber y Pei (2012, p. 130) y Hernández y Ferri (2006, p. 16).

Los DW pueden considerarse parte del proceso de inteligencia empresarial o de negocios (del inglés *Business Intelligence*, en adelante BI), un conjunto de actividades que tienen como fin el dar fundamentos para la toma de decisiones, de manera que éstas orienten acciones apropiadas y eficaces, de acuerdo con los objetivos fijados por la organización. En principio, la BI era aplicable únicamente a pequeñas y medianas empresas que buscaban aumentar su rentabilidad y disminuir costos, así como obtener ventajas competitivas en el mercado. Su implementación y desarrollo eran muy costosos. Sin embargo, actualmente se han desarrollado varias aplicaciones de BI con licencias de software gratuito y de código abierto (u *Open Source*), lo cual, aunado a la proliferación en la red de documentos, guías, comunidades, foros, tutoriales, cursos y wikis, han permitido ampliar el campo de aplicación de la BI a organizaciones más diversas. “Esta evolución permitió que el BI pueda ser aplicado a una biblioteca popular<sup>14</sup>, a un centro vecinal o cualquier institución u organización que necesite tomar decisiones pertinentes [con] relación a los datos que produce” (Bernabeu y García, 2017, p. 19).

Como es posible observar, los DW, al ser escindidos de los sistemas transaccionales, permiten adaptar y mejorar los datos para su posterior tratamiento. Si bien los DW existen desde la década de los ochenta, entre los desarrollos más significativos de los últimos diez años está la capacidad de almacenar datos semiestructurados y no estructurados junto con la forma tabular tradicional (Raina y Krishnamurthy, 2022, p. 168). La arquitectura y herramientas proveídas por los DW siguen constituyendo una importante plataforma para el análisis de datos y el OLAP, y éste último a su vez, una base importante para la realización de procesos de minería de datos, tema del cual hablaremos a continuación.

---

<sup>14</sup> En el caso de la biblioteca vecinal, los autores proponen aplicar la inteligencia de negocios para resolver, por ejemplo, las siguientes preguntas: ¿Cuál es la temática más consultada? ¿Qué días hay mayor concurrencia y por qué? ¿Qué libros deben ser adquiridos? ¿Cuál es el rango etario que lee cómics? ¿Qué perfil tienen las personas que leen determinada temática? (Bernabeu y García, 2017, p. 20).

### 1.4.3 Minería de datos

Han, Kamber y Pei definen a la minería de datos (en adelante MD) como el proceso de descubrir patrones interesantes y conocimiento [a partir del análisis] de grandes cantidades de datos. Las fuentes de datos pueden incluir bases de datos, almacenes de datos, la Web, repositorios de información, [datos transaccionales], o datos que se transmiten de manera dinámica al sistema (Han, Kamber y Pei, 2012, p. 8).

La minería de datos puede aplicarse como un nuevo medio para ampliar el conocimiento prácticamente en todas las áreas del quehacer humano donde se acumule información, [ya sea] resolviendo problemas, ayudando a la toma de decisiones, permitiendo una mejor comprensión de los fenómenos, llenando vacíos o facilitando proyecciones históricas (Colle, 2002. p. 23)

Los principales objetivos de la minería de datos en la práctica [del descubrimiento de conocimiento] suelen ser la predicción y la descripción. La **predicción** implica el uso de algunas variables o campos en la base de datos para pronosticar valores desconocidos o futuros de otras variables de interés. La **descripción** se centra en encontrar patrones interpretables por humanos que describan los datos (Fayyad, Piatetsky-Shapiro y Smyth, 1996, p. 12. Traducción propia). Los objetivos de predicción y descripción se logran mediante la realización de tareas de minería de datos mostradas en la Tabla 2.

**Tabla 2. Tareas de minería de datos**

Tarea	Descripción de la tarea	Aplicaciones
<b>Clasificación (Classification)</b>	Consiste en aprender una función que mapea (clasifica) un elemento de datos en una de varias clases predefinidas.	Clasificación de tendencias, identificación automatizada de objetos de interés en bases de datos visuales, etc.
<b>Regresión (Regression)</b>	Involucra aprender una función que mapea un elemento de datos en una variable de predicción de valor real.	Predicción, estimación de probabilidades, etc.
<b>Agrupamiento (Clustering)</b>	Es una tarea descriptiva común en la que se busca identificar un conjunto finito de categorías o grupos para describir los datos. Las categorías pueden ser mutuamente excluyentes y exhaustivas o consistir en una representación	Descubrir poblaciones homogéneas de usuarios, identificar segmentos de consumidores, etc.



Tarea	Descripción de la tarea	Aplicaciones
<b>Resumen (Summarization)</b>	<p>más rica, como las categorías jerárquicas o superpuestas.</p> <p>Implica métodos para encontrar una descripción compacta para un subconjunto de datos. Las técnicas de resumen a menudo se aplican al análisis interactivo de datos exploratorios y a la generación automatizada de informes.</p>	<p>Un ejemplo simple sería tabular la media y desviaciones estándar para todos los campos. Los métodos más sofisticados incluyen la derivación de reglas de resumen, técnicas de visualización multivariante y el descubrimiento de relaciones funcionales entre variables.</p>
<b>Modelado de dependencia / Aprendizaje de reglas de asociación (Dependency modeling / Association rules learning)</b>	<p>Consiste en encontrar un modelo que describa dependencias significativas entre variables. Los modelos de dependencia existen en dos niveles:</p> <p><u>Nivel estructural:</u> Especifica (a menudo en forma gráfica) qué variables dependen localmente entre sí.</p> <p><u>Nivel cuantitativo:</u> Especifica las fortalezas de las dependencias utilizando una escala numérica.</p> <p>Por ejemplo: las redes de dependencia probabilística utilizan independencia condicional para especificar el aspecto estructural del modelo y las probabilidades o correlaciones para especificar las fortalezas de las dependencias.</p>	<p>Desarrollo de sistemas probabilísticos expertos a partir de bases de datos en el área médica, recuperación de información y modelado del genoma humano. Algoritmo ampliamente usado para la configuración de sistemas de recomendación</p>
<b>Detección de cambios y desviaciones (Change and deviation detection)</b>	<p>Se enfoca en descubrir los cambios más significativos en los datos de valores previamente medidos o valores normativos.</p>	<p>Detección y análisis de valores atípicos en el conjunto de datos bajo análisis.</p>

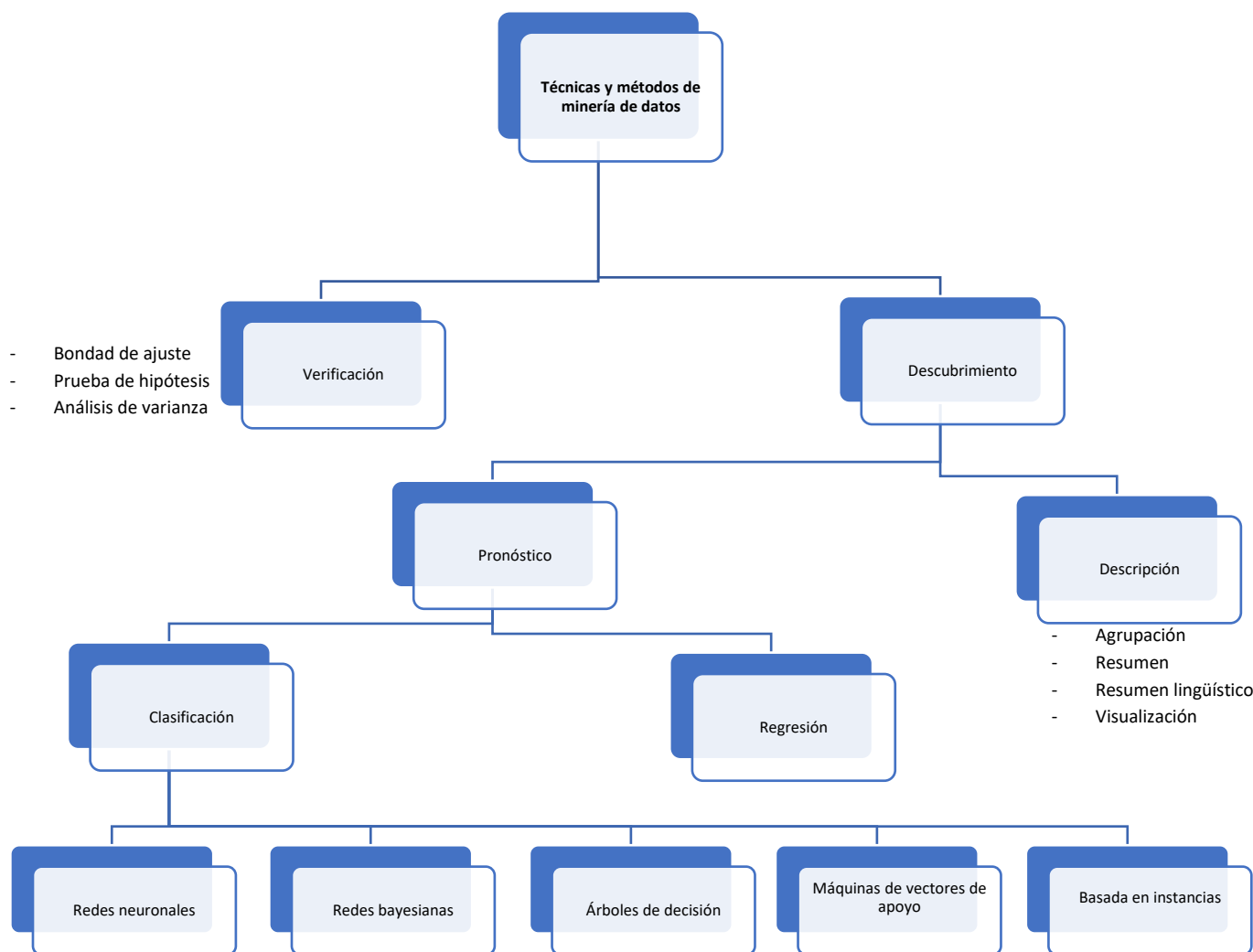
Fuente: Elaborada con información de Fayyad, Piatetsky-Shapiro y Smyth (1996, pp. 14-16). (Traducción propia).

Al aplicar la minería de datos para resolver alguna problemática dada, luego de determinar cuál o cuáles tareas son las que resultan de interés para los objetivos de tal aplicación, es necesario definir los algoritmos que ayudarán a cumplirlos. Estos algoritmos constituyen las técnicas o métodos de minería de datos<sup>15</sup>. De acuerdo a Fayyad et al. (1996), cada algoritmo de minería de datos tiene tres componentes principales: la **representación del modelo** (la manera en que describe los patrones detectables); la **evaluación del modelo** (el cálculo de qué tan bien un patrón -un modelo y sus parámetros- cumple con los criterios [de la metodología] de KDD); y la **búsqueda**

<sup>15</sup> Los términos de técnica, método y algoritmo son utilizados de manera indistinta por diversos autores del área de minería de datos. En esta tesis se considerarán como sinónimos, para señalar aquellas aplicaciones enfocadas a resolver las distintas tareas de la MD de acuerdo con la problemática a tratar.

(que incluye la indagación de parámetros que optimizan los criterios de evaluación del modelo y la detección del propio modelo en el conjunto de datos bajo análisis). Existe una amplia variedad de técnicas de minería de datos: árboles y reglas de decisión; regresión no lineal y métodos de clasificación; métodos basados en ejemplos, modelos probabilísticos de dependencia gráfica; modelos de aprendizaje relacional (Fayyad, Piatetsky-Shapiro y Smyth, 1996, pp. 17-22), entre otros.

Maimon y Rokach (2010) señalan que hay muchas técnicas de minería de datos utilizadas para diferentes propósitos y objetivos, por lo cual proponen una taxonomía que ayuda a comprender esa variedad de métodos, su interrelación y agrupación (Véase Figura 3). En ella distinguen, de inicio, dos tipos principales de técnicas de minería de datos: la orientada a la verificación (donde el sistema verifica la hipótesis del usuario) y la orientada al descubrimiento (en la cual el sistema encuentra nuevas reglas y patrones de manera autónoma).



**Figura 3. Taxonomía de técnicas y métodos de minería de datos**

Fuente: Maimon y Rokach (2010, p. 6).

Las técnicas de verificación se dedican a la evaluación de hipótesis propuestas por fuentes externas. Estos métodos incluyen las aplicaciones más comunes de las estadísticas tradicionales, las cuales están menos asociados a la minería de datos que sus contrapartes orientadas al descubrimiento. De éstas últimas, las técnicas orientadas a la predicción se enfocan en construir de manera automática modelos de comportamiento, que obtienen de muestras nuevas e indetectables a simple vista y pueden pronosticar valores de una o más variables relacionadas con la muestra.

También desarrollan patrones, los cuales reflejan el conocimiento obtenido de manera más comprensible y fácil de operar. Por otro lado, las técnicas descriptivas están orientadas a la interpretación de datos, que se centra en comprender (a partir de la visualización, por ejemplo) la forma en que los datos subyacentes se relacionan con sus partes.

Lo expuesto hasta este punto muestra parte de la gran variedad de tareas y técnicas que ofrece la minería de datos para la resolución de problemas. La aplicación de estas metodologías también ha permeado en el ámbito de las organizaciones, unidades o actividades de información, y en específico, de las bibliotecas, como se evidenciará más adelante. Antes de pasar a ello, se mostrará a continuación cuáles son los roles típicos de los participantes en proyectos que involucran la aplicación de la Ciencia de datos en organizaciones de diversa índole, incluidas las bibliotecas.

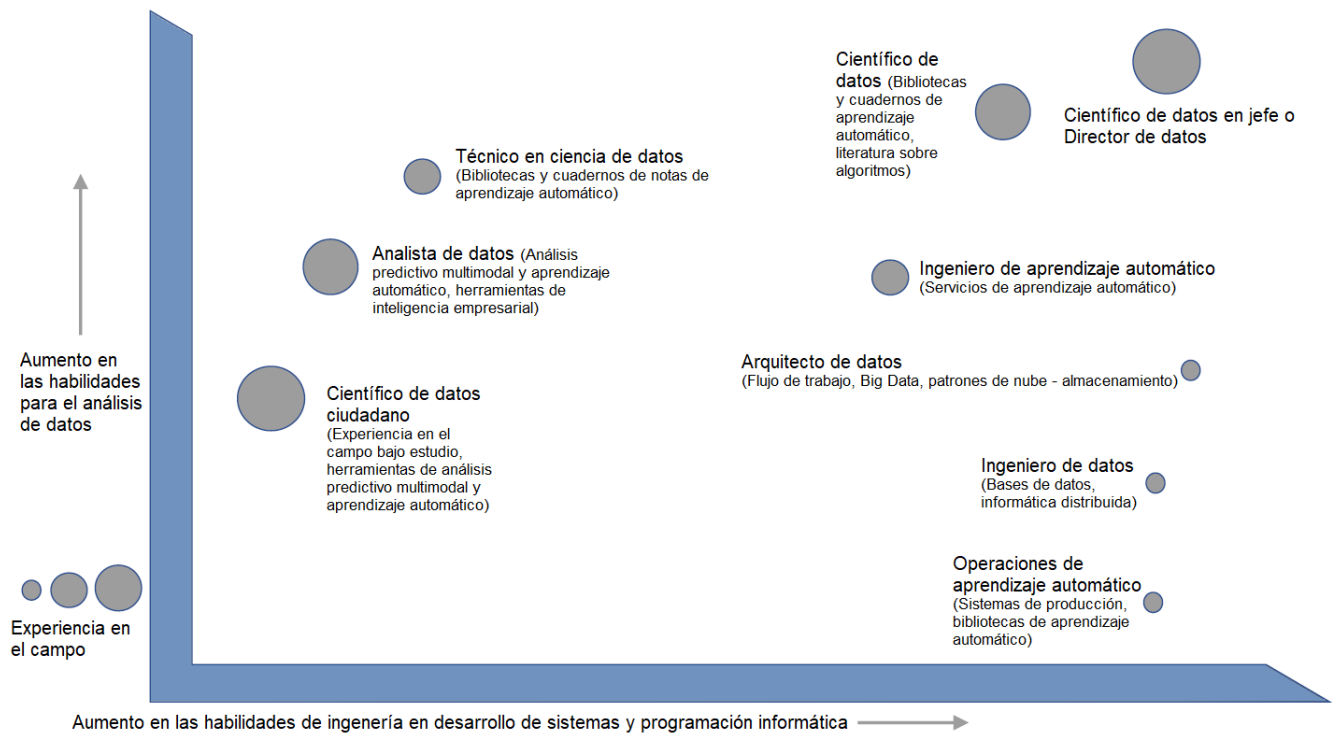
#### **1.4.4 Habilidades y roles de los participantes en la Ciencia de datos**

De acuerdo con Raina y Krishnamurthy (2022, p. 306-307), existen tres dimensiones de habilidades que son necesarias para desarrollar un proyecto de Ciencia de datos:

1. **Análisis de datos**: Es la capacidad de obtener información [útil] a partir de los datos, habilidad que puede aplicarse desde la interpretación de reportes estadísticos simples hasta la comprensión y ejecución de algoritmos de aprendizaje automático (ML, por las siglas en inglés de *Machine Learning*).
2. **Ingeniería en desarrollo de sistemas y programación**: Esta dimensión es un vasto campo de conocimientos en sí misma porque incluye varias disciplinas y especialidades. En el contexto de la Ciencia de datos, son necesarias habilidades de ingeniería para codificar los diversos pasos del proceso. Algunos aspectos relevantes incluyen:

- Diseño e implementación eficiente de algoritmos analíticos.
  - Preparación de datos: recopilación, limpieza y preparación de conjuntos de datos.
  - Conocimiento sobre la configuración de Unidades de Procesamiento de Gráficos (GPUs, por sus siglas en inglés de *Graphic processing units*) y de algoritmos de aprendizaje profundo, así como entrenamiento de modelos eficiente.
  - Producción de modelos de aprendizaje automático (ML).
3. **Experiencia en el campo**: La comprensión profunda del dominio [bajo estudio] y cómo los datos se relacionan con los procesos fundamentales del mismo, son esenciales para formular adecuadamente el problema a resolver, determinar el enfoque a emplear para llegar a una solución y también para evaluar que tan apropiada fue dicha solución.

En la Figura 4 se muestran los roles más comunes a identificar en un proyecto de Ciencia de datos, los cuales llegan a presentar en sus características una combinación de las habilidades antes enunciadas, dependiendo de las tareas que realizan cada uno de ellos.



**Figura 4. Roles de los participantes en equipos de Ciencia de datos y requerimientos de habilidades**

Fuente: Raina y Krishnamurthy (2022, pp. 306). (Traducción propia).

La determinación del equipo que participará en el desarrollo e implementación de proyectos de Ciencia de datos es una tarea de sumo interés si es que se desea llevar a buen término los emprendimientos. La diversidad de tareas a efectuar requiere de ciertas habilidades y conocimientos que generalmente es difícil encontrar en un solo individuo. Esto sin contar que una apropiada aplicación de Ciencia de datos implica la configuración de un grupo interdisciplinario de personas que aporten a la consecución de los objetivos planteados en el proyecto de análisis. Por tanto, existe la necesidad de que sus integrantes desarrollen habilidades de comunicación y de trabajo en ambientes interdisciplinarios que les permitan llevar a buen fin sus emprendimientos. En el siguiente apartado se llevará a cabo un recorrido por los proyectos identificados durante el desarrollo de esta tesis que constituyen ejemplos de aplicación de DW, DM y/o bibliominería en actividades bibliotecarias y de la información.

## **1.5 Aplicaciones de data warehouse y minería de datos en bibliotecas**

A continuación, se ofrece una recapitulación de los estudios que están enfocados a la mejora de procesos en bibliotecas y que en su gran mayoría han empleado principios de DW y MD para dicha labor. Dada la necesidad de optimizar la gestión y los resultados en bibliotecas, desde hace décadas se han llevado a cabo esfuerzos innumerables para contar con modelos y herramientas que permitan facilitar el proceso de evaluación y de toma de decisiones. Uno de ellos es el proyecto DECIMAL (Fisher y Oulton, 1995), una iniciativa británica cuyo objetivo fue identificar los procesos de toma de decisiones de pequeñas y medianas bibliotecas y determinar las necesidades de información de sus administradores, para así diseñar y desarrollar un módulo de apoyo para la toma de decisiones que fuese compatible con varios sistemas de gestión de bibliotecas. De acuerdo con los autores, es importante que los administradores de bibliotecas, como profesionales [de la información], conozcan mejor como recopilar y utilizar la información, estén al tanto de las capacidades de otros sistemas de información que no son del uso exclusivo de las bibliotecas [pero que pueden ser aplicados en éstas] y comprendan mejor sus propios procesos de toma de decisiones (Fisher y Oulton, 1995, p.18). A pesar de que este proyecto en particular no hace referencia a la MD aplicada a bibliotecas, se considera importante mencionarlo, pues una de las fases más importantes de la MD y el diseño de DW es el determinar cabalmente las necesidades de información de los tomadores de decisiones y este artículo hace aportaciones importantes en este rubro.

Banerjee (1998) fue uno de los primeros en dar a conocer cuáles podrían ser las aplicaciones de MD en bibliotecas, describió las funciones de MD y en qué áreas podrían implementarse, además de señalar los problemas que su puesta en marcha podría enfrentar y los requisitos previos necesarios a considerar por los centros de información antes de proceder a su uso. Banerjee brinda ejemplos teóricos del empleo de MD en bibliotecas en tareas de búsqueda y recuperación de información y la distingue de otras tecnologías de búsqueda en cuanto a que permite realizar elecciones y cálculos para los buscadores a partir del uso de algoritmos, para luego categorizar la información tomando en cuenta tales elecciones (Banerjee,1998).

También a finales de la década de los noventa, Scott Nicholson comienza su trabajo en el tema de la minería de datos aplicada a bibliotecas con una tesis doctoral en la cual presentó un modelo estadístico para crear un agente de información que discrimina entre páginas web que albergan trabajos de investigación académica de otras páginas web (Nicholson, 2000). Para ello, llevó a cabo un análisis de páginas web empleando un conjunto de criterios previamente delimitados y posteriormente empleó un árbol de clasificación para llegar a una decisión con respecto a qué tipo de página se tiene bajo la mira, permitiendo así su clasificación. Este fue solo el comienzo de una serie de aportaciones por parte de Nicholson que promoverían la difusión del uso de la MD en bibliotecas, así como la búsqueda de nuevas aplicaciones en otros ámbitos de los centros de información.

Kim Guenther, bibliotecóloga estadounidense involucrada en el tema de desarrollo Web, ofreció una aproximación teórica al uso de la MD en bibliotecas y para ello determinó tres etapas: selección y adquisición de datos; preparación y procesamiento e interpretación e integración (Guenther, 2000). Para Guenther, si bien la MD no se realizaba en la mayoría de las bibliotecas de manera sofisticada, sus conceptos centrales podrían aplicarse para ayudarnos a adaptar los servicios para satisfacer las necesidades de los usuarios en constante evolución.

Una implementación muy completa de la MD y DW se presenta en la tesis de Ashwin Needamangala (2000)<sup>16</sup>, quien elabora una propuesta de diseño e implementación de un DW, así como la aplicación de la técnica de redes neurales de MD en los datos almacenados, todo ello para constituir un sistema de apoyo para la toma de decisiones de la *University of Florida Library* y permitirles realizar análisis financieros que efficienten el uso del presupuesto, además de apoyar a la selección de bibliografía. En este caso, la MD fue empleada en el análisis de registros que pertenecen a una colección física.

Por su parte, Bollen y Luce emplearon en 2002 una técnica de análisis jerárquico de clúster para diseñar una metodología que, a partir del estudio de los patrones de

---

<sup>16</sup> Véase también Su, Siew-Phek T. y Ashwin Needamangala. (2000). Harvesting information from a library data warehouse. *Information Technology in Libraries*, 19(1), p. 17-28.



recuperación de información de los usuarios en una biblioteca digital (los cuales son reflejados por los registros del servidor), permitió evaluar el impacto de la colección y determinar la estructura de la comunidad de usuarios (Bollen y Luce, 2002).

Kao, Chang y Lin (2003) desarrollaron a partir de la MD un modelo al que denominaron ABAMDM, cuyo propósito fue ofrecer apoyo para la toma de decisiones relacionadas a la asignación presupuestal para la adquisición de bibliografía en una biblioteca académica considerando los datos históricos de circulación de materiales, con lo cual se posibilitó el otorgar un mayor presupuesto a los departamentos académicos que hacían más uso de los materiales documentales.

También en 2003, Scott Nicholson retoma el tema de la minería de datos aplicada a bibliotecas y, junto a Jeffrey Stanton, comienza a utilizar el término de bibliominería (Nicholson y Stanton, 2003). Dentro de las aportaciones de Nicholson durante este periodo figuran las siguientes: elabora un artículo en el cual explica la metodología empleada en un proceso de bibliominería y brinda un enfoque especial a la creación del DW y a la protección de la privacidad de los usuarios (Nicholson, 2003a), retoma su tesis doctoral con un artículo en el cual propone la creación de un agente inteligente para la identificación y clasificación de contenido académico en Internet utilizando el programa Perl para analizar páginas web (considerando una lista de 41 criterios previamente definidos por el autor para evaluar la calidad del contenido académico encontrado). La bibliominería fue aplicada posteriormente al conjunto de datos recolectados con el fin de clasificar los contenidos y crear así una biblioteca digital con fuentes académicas de calidad (Nicholson, 2003b).

Las bibliotecas digitales han sido de particular interés para los investigadores que buscan aplicar la bibliominería. Por ejemplo, en Grecia analizaron el comportamiento de los usuarios a partir de las consultas hechas por éstos a una biblioteca digital y, utilizando técnicas de minería de datos, pudieron delimitar comunidades de usuarios con intereses comunes (Papatheodorou et. al, 2003).

Nicholson siguió su trabajo en el tema de evaluación bibliotecaria, en 2004 presentó un “marco conceptual para la medición holística y la evaluación acumulativa de los servicios

bibliotecarios” fundamentado en la teoría de sistemas que tuvo buena acogida en el ámbito profesional, pues, como veremos más adelante, fue retomada posteriormente tanto por él como por otros autores como base para realizar varios estudios<sup>17</sup>. Inclusive propuso, considerando las tareas de la biblioteconomía basada en evidencias (EBL), llevar a cabo tareas de arqueología en bibliotecas digitales para comprender la interacción entre usuarios y recursos informativos a partir de la aplicación de un marco conceptual para comprender el uso de la biblioteca a partir de la evaluación basada en artefactos y del uso de la bibliominería y el DW (Nicholson, 2005). La EBL, así como los temas de la integración de la minería de datos a los servicios de biblioteca digital y los relacionados con la privacidad de los usuarios de las bibliotecas continuaron siendo un tema recurrente en las publicaciones de Nicholson<sup>18</sup>, que de esta manera se consolida como referente en procesos de bibliominería.

Conforme pasa el tiempo y debido a la tendencia al alza del uso de materiales en formato electrónico, la MD aplicada a bibliotecas digitales sigue cobrando auge: Chang y Chen (2006) emplearon tecnología de MD para descubrir ciertos grupos de lectores en registros de préstamos pasados en una biblioteca digital, ello con el fin de clasificarlos y conocer sus características. Sus conclusiones mostraron que, dependiendo de los resultados de MD al categorizar a los usuarios, era posible que una biblioteca anticipara las necesidades de información de sus lectores.

En la búsqueda de literatura fue posible ubicar la tesis de Gutiérrez y Barranco (2008), quienes aplicaron exitosamente el KDD y la MD para realizar un análisis de la circulación de materiales en la Biblioteca Central de la Universidad Nacional Autónoma de México empleando los programas informáticos Access y Weka. De esta manera pudieron delimitar el comportamiento de los usuarios tomando como variable principal la carrera asignada en el sistema, ello con la finalidad de obtener tendencias respecto al préstamo y devolución del material de la biblioteca y apoyar el proceso de TD del Departamento

---

<sup>17</sup> Véase Anexo 1. Resumen de la matriz de medición y evaluación holística de Nicholson. En tal propuesta, el autor determinó una matriz de evaluación bidimensional en la cual las columnas representan el tema a evaluar (ya sea el sistema de biblioteca o su uso), mientras que las filas indican la perspectiva bajo la cual se desarrolla la evaluación (que puede ser interna o externa) (Siguenza-Guzman, et al. 2015, p. 7).

<sup>18</sup> Véanse: Nicholson (2006a), Nicholson (2006b), Nicholson (2006c), Nicholson y Stanton (2006) y Nicholson y Arnott (2007).

de Circulación Bibliográfica de la Biblioteca Central. Las autoras hicieron énfasis en incluir un análisis del negocio y sus necesidades previo a la implementación de un proceso de KDD, e indicaron que para ello también “convendría incluir la construcción de un *data warehouse* que facilite la preparación, limpieza y por ende el análisis de los datos, mejorando su calidad” (Gutiérrez y Barranco, 2008, p. 135).

En 2012, Cox y Jantti dieron a conocer una aplicación práctica de DW en bibliotecas, a partir de la creación de un cubo de biblioteca consistente en un DW multidimensional que une los datos de uso de la biblioteca con los datos demográficos de los estudiantes y sus calificaciones promedio ponderadas. El propósito del proyecto reside en demostrar el valor de la biblioteca de la *Wollongong University* para apoyar la enseñanza y el aprendizaje, así como recopilar información para ayudar a la biblioteca a mejorar el impacto de sus recursos y actividades de enseñanza con respecto al rendimiento académico y la participación de los estudiantes (Cox y Jantti, 2012, p. 308).

En el ámbito Latinoamericano surgió el proyecto colombiano Elogim, una implementación del KDD y la bibliominería utilizados para la elaboración de informes estadísticos sobre el uso de los recursos electrónicos (empleando la aplicación EZproxy<sup>19</sup>), labor que aporta información relevante para la toma de decisiones y la gestión de unidades de información académicas (Quiroz y Valencia, 2012). Elogim fue desarrollado por la Biblioteca Fundadores de la Universidad CES de Colombia con el propósito de definir indicadores de gestión, definir patrones de comportamiento de los usuarios de recursos electrónicos, innovar los servicios y actualizar los procesos. Hoy en día continua vigente ofreciendo servicios de inteligencia de negocios a bibliotecas.

También en Latinoamérica, específicamente en Ecuador, fue presentada la tesis de Haro y Pérez (2014) quienes desarrollaron un *Data warehouse para el Centro de Documentación Regional “Juan Bautista Vázquez”*. Su estudio presenta el análisis, diseño e implementación de un DW de manera muy detallada empleando la metodología

---

<sup>19</sup> EZproxy es un producto de autenticación y acceso a recursos electrónicos de OCLC que “permite a las bibliotecas proporcionar recursos electrónicos a los usuarios de manera simple y segura sin importar dónde y cuándo realicen búsquedas” (OCLC, 2020).

Hefesto (Bernabeu y García, 2017), así como la aplicación de técnicas de bibliominería empleando las herramientas informáticas de código abierto Pentaho BI y Weka.

Otra importante aportación de DW y MD aplicada a bibliotecas es la tesis doctoral presentada en Bélgica por la ecuatoriana Lorena Catalina Siguenza-Guzman (2015). La autora elaboró un detallado estudio en el cual diseñó e implementó un DW a partir del cual se visualizan y generan informes de datos para que los administradores de bibliotecas publiquen indicadores de una manera simple y rápida mediante el uso de herramientas de informes en línea. Siguenza-Guzman también examinó tres técnicas de minerías de datos: regresión, agrupamiento y clasificación; las cuales aplicó a su estudio de caso para predecir la inversión futura en el desarrollo de bibliotecas, encontrar grupos de usuarios que comparten intereses comunes y perfiles similares pero que pertenecen a diferentes facultades, así como para predecir los factores de la biblioteca que afectan el rendimiento académico de los estudiantes mediante el análisis de las correlaciones entre uso del centro de información y su desempeño escolar.

En cuanto al análisis del estado del arte en aplicaciones de MD en bibliotecas resultó particularmente valioso el artículo publicado por Lorena Siguenza-Guzman, Víctor Saquicela, Elina Ávila-Ordóñez, Joos Vandewalle y Dirk Cattrysse, quienes llevaron a cabo una revisión de la literatura sobre la aplicación de técnicas de minería de datos en bibliotecas en publicaciones hechas durante el periodo 1998-2014 (Siguenza-Guzman, et al. 2015). Los autores identificaron, analizaron y clasificaron 41 artículos empleando un marco de análisis teórico fundamentado en la propuesta para la evaluación holística de bibliotecas hecha por Nicholson (2004), que, a decir de los autores, constituye uno de los enfoques más completos para evaluar bibliotecas desde una perspectiva integral.

Ben Showers incluyó en la obra *Library analytics and metrics* (2015, p. 28) un estudio de caso donde se detallan las características de *Haystacks*, un kit de herramientas analíticas para bibliotecas que fue desarrollado por el *Harvard Library Innovation Lab*. Su objetivo: a partir de un panel de control de gran navegabilidad permitir la visualización avanzada de múltiples tipos de datos referentes al trabajo desarrollado en las bibliotecas de la Universidad de Harvard, de manera tal que, tanto usuarios como personal de la biblioteca, puedan obtener información óptima para mejorar su toma de decisiones. En

este caso, fue patente la necesidad de contar con un desarrollador interno y al menos tres desarrolladores externos al equipo del *Harvard Library Lab* para llevar a buen puerto el proyecto.

Ruiz y Romero (2017) aplicaron algoritmos estadísticos a una base de datos bibliográfica con el fin de buscar patrones ocultos. Su estudio “encontró patrones a partir de la clusterización o agrupamiento con los cuales fue posible potenciar los servicios de difusión selectiva de la información y proponer nuevos servicios para que formen parte de los productos que brinda la biblioteca” (Ruiz y Romero, 2017, p. 209).

En el campo del diseño de DW concerniente al área bibliotecológica fue posible ubicar la aportación de Gorbea y Madera (2017), quienes propusieron el uso de un DW “orientado a la medición del desarrollo disciplinar en las ciencias bibliotecológica y de la información en instituciones académicas de Iberoamérica”. Los resultados obtenidos por este estudio “permiten medir el comportamiento métrico del desarrollo disciplinar en la temática y región de referencia por instituciones y países, además de establecer tendencias que permitan aportar información para la toma de decisiones en materia de investigación, docencia y desarrollo disciplinar” (Gorbea y Madera, 2017, p. 166-167).

Durante la búsqueda de propuestas de aplicación de MD en bibliotecas fue posible encontrar que en la Escuela de Bibliotecología y Ciencias de la Información de la Universidad de Costa Rica existe desde 2016 un proyecto de investigación vigente denominado “Bibliominería de datos en el campo de las Ciencias Bibliotecológicas y de la Información”<sup>20</sup>, cuyos responsables son Mynor Fernández Morales, Róger Bonilla Carrión, José Iván Saborío Acuña y Ricardo Chinchilla Arley. El objetivo general del proyecto es “investigar las posibilidades que ofrece la minería de datos aplicada a las bibliotecas, conocida como bibliominería, con la finalidad de estudiar los métodos para extraer información que se encuentra en forma oculta en los grandes volúmenes de almacenamiento que poseen las unidades de información, permitiendo la realización de análisis de las proyecciones y tendencias que existen sobre los datos almacenados, que

---

<sup>20</sup> Véase Fernández y Bonilla (2020).

ayuden a facilitar la toma de decisiones en las unidades para una mejor gestión de los recursos disponibles” (Sandi, 2019, p. 10).

En la revisión de literatura realizada para este proyecto se ubicó también el aporte del ingeniero mexicano Marcial Contreras (2018)<sup>21</sup>, quien en su tesis doctoral del área bibliotecológica propuso un modelo de minería de texto que apoya la labor de organizar la información documental. En su caso, Contreras aplicó minería de texto a los registros de la base de datos del catálogo LIBRUNAM “con la finalidad de desarrollar un sistema de cómputo con la capacidad de realizar la clasificación de documentos de acuerdo [con] la clasificación de la Biblioteca del Congreso (*Library of Congress* o LC) y su posterior aplicación en la asignación temática de la información documental” (Contreras 2018, p. 120).

En lo que respecta a servicios de información comercial especializados para la toma de decisiones de bibliotecas que aplican metodologías de DW, MD y/o bibliominería, se encontró que actualmente existen varias propuestas y proveedores en el mercado<sup>22</sup>. Sin embargo, dado que son servicios externos o adicionales a los ofrecidos en plataformas de gestión bibliotecaria, es necesario que los centros de información paguen ciertas cuotas para obtener tales servicios, lo cual en la mayoría de las bibliotecas universitarias públicas resulta prohibitivo por la falta de presupuesto. Lo anterior ratifica la necesidad de dar a conocer el *know-how*<sup>23</sup> que subyace a las implementaciones de DW y MD aplicadas a bibliotecas, de manera que, aquellas interesadas en su empleo sepan cómo llevarlas a cabo en caso necesario.

---

<sup>21</sup> Véase también Contreras (2014).

<sup>22</sup> Véase Anexo 2. Proveedores de servicios de información especializada para la toma de decisiones en bibliotecas.

<sup>23</sup> Saber hacer (Trad.).

# CAPÍTULO 2. Metodología

La elección de la Biblioteca de la Dependencia de Educación Superior de Salud (en adelante BDESSalud) del Sistema Universitario de Bibliotecas Académicas de la Universidad Autónoma de Chihuahua como unidad de análisis fue motivada por el trabajo previo realizado por la autora en su tesis de maestría, en la cual delimitó un modelo de evaluación aplicable a bibliotecas universitarias, así como el deseo de brindar a su Alma Mater, una universidad pública estatal, nuevas metodologías y herramientas que apoyen a su apropiada gestión. Asimismo, la elección estribó en que la BDESSalud puede ser tomada como un ejemplo típico de biblioteca universitaria en México, por lo cual se considera que emplearla como unidad de análisis puede motivar la aplicación del modelo propuesto en otros sistemas similares.

La finalidad del presente capítulo es describir los aspectos metodológicos considerados para lograr los objetivos planteados en esta tesis. Para ello se comienza con una descripción de la unidad de análisis y observación, la enumeración de los componentes del modelo de integración de datos en el DW y el procedimiento de trabajo llevado a cabo con los datos para llegar a la aplicación de bibliominería.

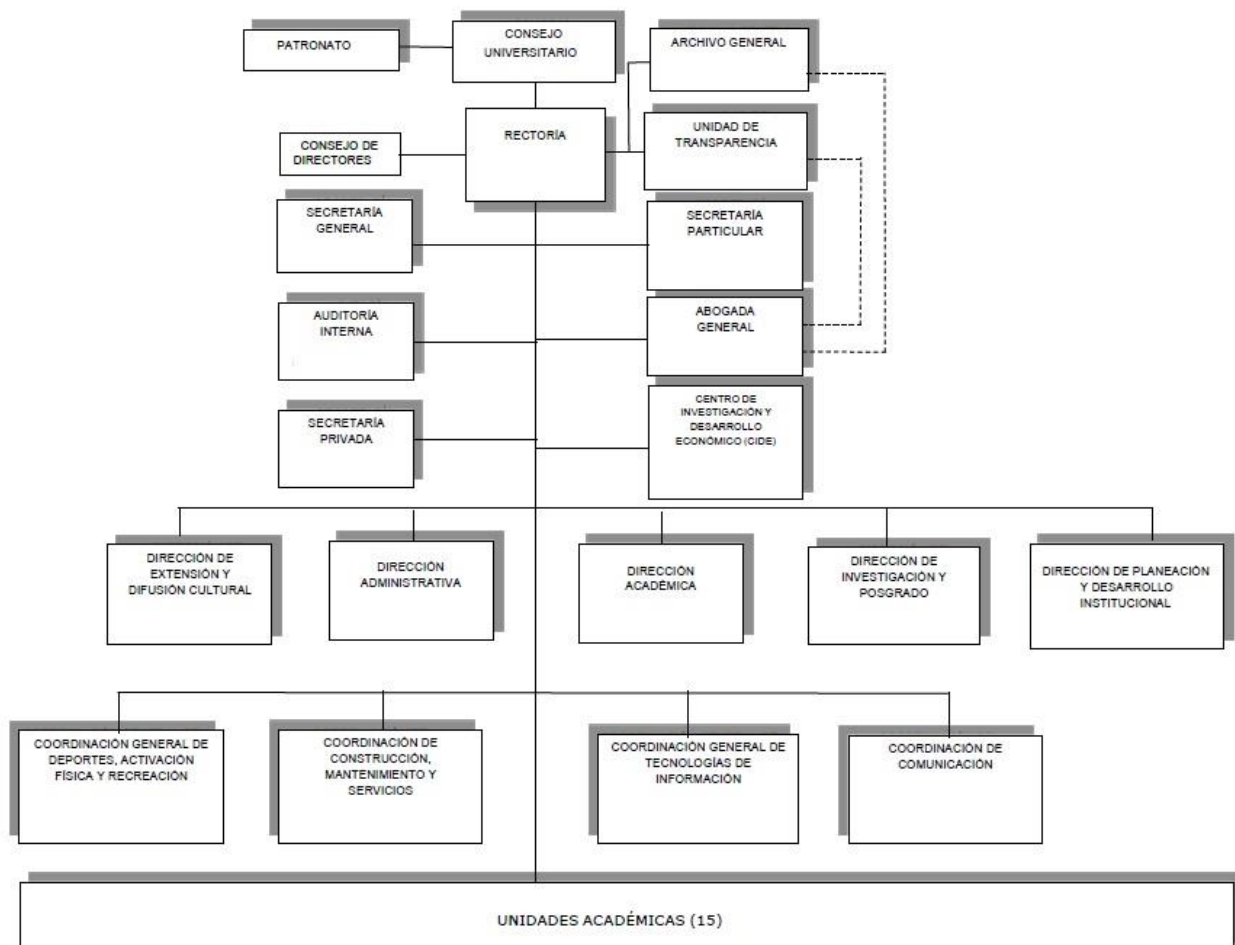
## 2.1 Unidad de análisis y observación

### 2.1.1 Biblioteca de la DES de Salud del SUBA, UACH

Para dar cuenta de la unidad de análisis y observación considerada para la realización de esta tesis es necesario conocer el contexto que la enmarca. Así, se comienza por indicar que la BDESSalud pertenece a la red de bibliotecas que alberga el Sistema Universitario de Bibliotecas Académicas (en adelante SUBA) y que la institución educativa de la cual depende tal sistema bibliotecario es la Universidad Autónoma de

Chihuahua (en adelante UACH). Esta institución educativa (de carácter público estatal) fue creada por decreto gubernamental del entonces gobernador del estado de Chihuahua, C. Oscar Soto Máñez, el 08 de diciembre de 1954. En su *Estadística básica 2020* (Véase Universidad Autónoma de Chihuahua, 2020b), la UACH declaró contar con una matrícula total de 29,573 estudiantes, de los cuales 26,753 pertenecían a licenciatura; 1,337 al posgrado; 166 a técnicos en educación media superior (especialidad enfermería) y 1,337 al sistema abierto y virtual. El total de docentes contabilizados fue de 3,029, de los cuales 899 eran de tiempo completo; 113 de medio tiempo y 2,017 de hora clase. Para entonces, la UACH contaba con 14 unidades académicas (facultades) y un centro universitario (Parral) cuya oferta educativa en conjunto era de 136 programas: 63 de licenciatura; 71 de posgrado y dos de técnico nivel medio superior. Veintiuno de estos programas estaban incluidos en el Padrón Nacional de Posgrado de Calidad (PNPC) del Consejo Nacional de Ciencia y Tecnología (CONACYT) (UACH, 2020b, *Estadística básica*, p. 7). La Figura 5 muestra el organigrama de la UACH donde se resume la estructura que le permite su funcionamiento.



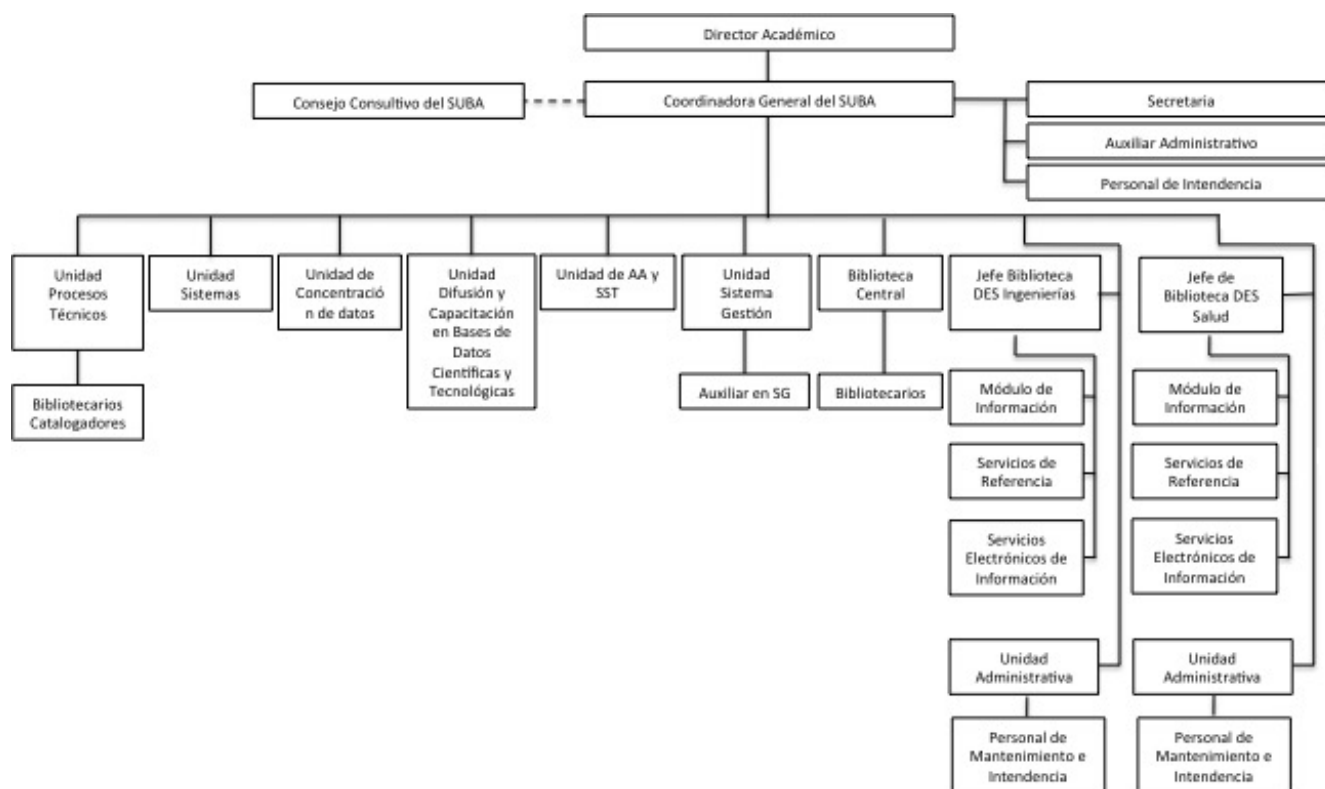


**Figura 5. Organigrama de la Universidad Autónoma de Chihuahua**

Fuente: *Organigrama general UACH* (Universidad Autónoma de Chihuahua, 2017a)

En el esquema aún no es evidente la existencia del sistema de bibliotecas. Sin embargo, aquí se acotará que el SUBA depende directamente de la Dirección Académica de la UACH, lo cual tiene implicaciones importantes en la manera en que desempeña sus funciones. Considerando la historia particular del SUBA, su creación fue aprobada por el H. Consejo Universitario de la UACH el 14 de abril del 2000, como consta en el Acta de Consejo número 382 (Aceves, 2002). A partir de ese momento se tuvo el reto de unificar la operación de las bibliotecas, por lo cual se comenzó a trabajar con las bases de datos que contenían los registros de la totalidad de los materiales albergados por estos centros de información, que fueron integrados empleando el sistema de automatización Logicat.

A la par se realizaron adquisiciones de mobiliario especializado, equipo de cómputo y recursos electrónicos de información necesarios para la prestación de los servicios. La organización del SUBA se muestra gráficamente en la Figura 6.



**Figura 6. Organigrama del Sistema Universitario de Bibliotecas Académicas (SUBA)**

Fuente: *Organigrama SUBA* (Sistema Universitario de Bibliotecas Académicas, 2017)

El Consejo Consultivo del SUBA (en adelante CCSUBA) se integra por la Coordinación General del SUBA, los jefes de unidad de la Coordinación y los jefes de cada una de las bibliotecas que conforman el sistema. Para el 2020 el SUBA tenía incorporadas 18 bibliotecas, de las cuales la Biblioteca Central y las Bibliotecas de las Dependencias de Educación Superior (DES) de Ingenierías y de Salud están subordinadas a la Coordinación del SUBA. El resto de ellas depende de cada unidad académica o centro

universitario al que están adscritas. En la Tabla 3 se muestra el listado de las bibliotecas del SUBA.

**Tabla 3. Listado de bibliotecas que conforman el SUBA**

1	Biblioteca Central - Odontología
2	Biblioteca de la DES de Ingenierías
3	Biblioteca de la DES de Salud
4	Biblioteca del Centro Universitario Parral
5	Biblioteca de la Facultad de Artes
6	Biblioteca “Eva Mariscal Robles”, Facultad de Ciencias Agrotecnológicas
7	Biblioteca de Facultad de Ciencias Agrotecnológicas, Extensión Cuauhtémoc
8	Biblioteca “Reyes Humberto de las Casas”, Facultad de Ciencias Agrícolas y Forestales
9	Biblioteca “José Fuentes Mares”, Facultad de Ciencias Políticas y Sociales, Ciudad Juárez
10	Biblioteca de Facultad de Ciencias Políticas y Sociales, Chihuahua
11	Biblioteca “Prof. Gabriel Rodríguez Arzate”, Facultad de Ciencias de la Cultura Física
12	Biblioteca “C. P. Antonio Horcasitas Barrio”, Facultad de Contaduría y Administración
13	Biblioteca de Facultad de Contaduría y Administración, Extensión Camargo
14	Biblioteca de Facultad de Contaduría y Administración, Extensión Delicias
15	Biblioteca “Oscar Ornelas K.”, Facultad de Derecho
16	Biblioteca “Bertrand Russell”, Facultad Académica de Filosofía y Letras
17	Biblioteca de Medicina y Ciencias Biomédicas, Extensión Colón
18	Biblioteca “M. V. Federico Rubio Lozano”, Facultad de Zootecnia y Ecología

Fuente: Directorio de Bibliotecas (Sistema Universitario de Bibliotecas Académicas, 2021a)

La unidad de análisis y observación de esta tesis es la BDESSalud, centro de información que resultó de la unión de los acervos de las bibliotecas de la Facultad de Enfermería y Nutriología (en adelante FEN) y la Facultad de Medicina y Ciencias Biomédicas (en adelante FMCB) de la UACH en julio del 2010 (Terrazas Guzmán, 2018). La BDESSalud brinda sus servicios a toda la comunidad de la UACH<sup>24</sup> (inclusive a usuarios externos), empero, la población de usuarios pertenecientes a estas dos unidades académicas es a la que atiende de manera más directa. Para 2021 los programas de estudio ofrecido en la FEN y la FMCB son los que se muestran en la Tabla 4.

**Tabla 4. Programas de estudio ofrecidos en las unidades académicas que atiende la BDESSalud**

<b>Nivel</b>	<b>Facultad de Enfermería y Nutriología</b>	<b>Facultad de Medicina y Ciencias Biomédicas</b>
<b>Técnico</b>	Enfermería General con Bachillerato (formación de técnicos en educación media superior con especialidad en enfermería)	No aplica
<b>Licenciatura</b>	Licenciatura en Enfermería Licenciatura en Nutrición	Licenciatura de Médico Cirujano y Partero (Campus Universitario II)  Licenciatura en Ingeniería Biomédica (Campus Universitario II)  Licenciatura en Terapia Física y Rehabilitación (Extensión de la Facultad de Medicina y Ciencias Biomédicas, Colón)
<b>Posgrado</b>	Maestría en Enfermería (pertenece al PNPC)  Maestría en Salud en el Trabajo (pertenece al PNPC)  Maestría en Nutrición Humana	<b>Especialidades</b>  <u>Quirúrgicas de Entrada Directa:</u> Cirugía General Ginecología y Obstetricia Traumatología y Ortopedia  <u>Quirúrgicas de Entrada Indirecta:</u> Angiología y Cirugía Vasculat Cirugía Plástica y Reconstructiva  <u>No Quirúrgicas:</u> Anestesiología Geriatría Medicina en el Enfermo en Estado Crítico Medicina Interna Pediatria Médica Radiología e Imagen Nefrología Medicina del Trabajo y Ambiental

<sup>24</sup> Sobre todo, a usuarios de las Unidades Académicas del área de las Ciencias de la Salud: Facultad de Enfermería y Nutriología, Facultad de Medicina y Ciencias Biomédicas, Odontología y Facultad de Ciencias de la Cultura Física.

<b>Nivel</b>	<b>Facultad de Enfermería y Nutriología</b>	<b>Facultad de Medicina y Ciencias Biomédicas</b>
		Urgencias Médico-Quirúrgicas  <u>Alta Especialidad:</u> Biología de la Reproducción Humana Cirugía Articular Cirugía Ginecológica Avanzada de Mínima Invasión Neumología Pediátrica  <u>Por Hospital:</u> Hospital Central Universitario "Dr. Jesús Enrique Grajeda" Hospital General "Dr. Salvador Zubirán Anchondo" Hospital Infantil de Especialidades de Chihuahua IMSS ISSSTE  <b>Maestría en Formación Biomédica (pertenece al PNPC)</b>  <b>Maestría en Dirección y Gestión en Salud</b>  <b>Maestría en Ciencias Biomédicas</b>

Fuentes: UACH, 2020b, Estadística básica (pp. 31, 35, 48, 51) y Facultad de Medicina y Ciencias Biomédicas (2021a, 2021b y 2021c).

En la Tabla 5 se muestran los programas de estudio de la FEN y la FMCB que han sido acreditados.

**Tabla 5. Programas de estudio acreditados**

<b>Unidad académica</b>	<b>Programa de estudio</b>	<b>Ente evaluador</b>
Facultad de Enfermería y Nutriología	Licenciatura en Enfermería	Comités Interinstitucionales para la Evaluación de la Educación Superior (en adelante CIEES) Consejo Mexicano para la Acreditación de Enfermería A. C. (COMACE), organismo reconocido por el Consejo para la Acreditación de la Educación Superior, A. C. (COPAES)
Facultad de Enfermería y Nutriología	Licenciatura en Nutrición	Consejo Nacional para la Calidad de Programas Educativos en Nutriología, A. C. (CONCAPREN), organismo reconocido por el COPAES
Facultad de Medicina y Ciencias Biomédicas	Licenciatura de Médico Cirujano y Partero	CIEES Consejo Mexicano para la Acreditación de la Educación Médica, A. C. (COMAEM), organismo reconocido por el COPAES
Facultad de Medicina y Ciencias Biomédicas	Ingeniería Biomédica	Consejo de Acreditación de la Enseñanza de la Ingeniería, A. C. (CACEI).

Fuente: Cuarto Informe Anual de Actividades UACH (Universidad Autónoma de Chihuahua, 2020a, p. 63 y 288).

Para agosto de 2020, la FEN y la FMCB contaban con los grupos disciplinares de Enfermería y Nutriología (FEN) y Medicina y Ciencias Biomédicas (FMCB). Tales grupos fueron creados como parte de una estrategia para dar seguimiento a la productividad de sus integrantes y evaluarlos constantemente para crear y fortalecer Cuerpos Académicos que pudiesen ser reconocidos por el Programa para el Desarrollo Profesional Docente para el Tipo Superior (PRODEP) de la Secretaría de Educación Pública (SEP)<sup>25</sup>. Para el año 2021, las unidades académicas a las cuales presta sus servicios la BDESSalud contaban con los cuerpos académicos reconocidos ante PRODEP indicados en la Tabla 6.

**Tabla 6. Cuerpos académicos de las unidades académicas atendidas por la BDESSalud**

<b>Unidad académica</b>	<b>Cuerpo Académico</b>	<b>Estatus</b>
Facultad de Medicina y Ciencias Biomédicas	UACH-CA-28: Salud, Trabajo y Ambiente	Consolidado
Facultad de Medicina y Ciencias Biomédicas	UACH-CA-19: Biomedicina Molecular	En consolidación
Facultad de Medicina y Ciencias Biomédicas	UACH-CA-133: Patogenia y Biomédica Molecular	En consolidación
Facultad de Enfermería y Nutriología	UACH-CA-102: Educación, Enfermería y Salud	En formación
Facultad de Enfermería y Nutriología	UACH-CA-23: Educación, Salud y Trabajo	En formación

Fuente: Universidad Autónoma de Chihuahua (2020a, p. 119 y p. 123).

La BDESSalud está sujeta a la misión, visión y objetivos fijados por el SUBA. La misión del sistema (y, por ende, de la BDESSalud) es “[a]poyar con espíritu de servicio y compromiso las funciones sustantivas de generación, transmisión y extensión del conocimiento de la UACH, a través de la organización y difusión de los recursos de información con tecnología de vanguardia y la formación de usuarios autosuficientes”. Su visión es que “[e]l SUBA será un organismo que administre las bibliotecas académicas, con capacidad para tomar decisiones sobre los recursos humanos,

<sup>25</sup> El objetivo de este programa es “profesionalizar a los Profesores de Tiempo Completo (PTC) para que alcancen las capacidades de investigación-docencia, desarrollo tecnológico e innovación y con responsabilidad social, se articulen y consoliden en cuerpos académicos y con ello generen una nueva comunidad académica capaz de transformar su entorno”. (Secretaría de Educación Pública, 2021).

informativos, tecnológicos y financieros; en un ambiente participativo, equitativo y armónico que apoye la formación de profesionales que contribuyan al desarrollo de la sociedad” (Sistema Universitario de Bibliotecas Académicas, 2021b). En cuanto a la normativa de operación establecida por el SUBA para el logro de su misión y visión<sup>26</sup>, la BDESSalud debe cumplir con los siguientes puntos:

- Satisfacer las necesidades de información de los usuarios, en cumplimiento con los fines de la Universidad y basados en los principios de libre acceso, legalidad, eficacia, eficiencia, transparencia y calidad en la prestación de servicios.
- Ofrecer programas de capacitación y actualización al personal [...], de forma sistemática y continua, compartiendo los resultados de sus investigaciones y experiencias.
- Trabajar en forma integrada con los miembros de la comunidad universitaria, en el cumplimiento de los fines esenciales de la Universidad.
- Evaluar de manera continua la colección documental para detectar fortalezas y debilidades, con el fin de contar con elementos para realizar una selección adecuada.
- Desarrollar y mantener una colección adecuada, suficiente y actualizada, que cubra [...] los programas académicos de nuestra institución.
- Administrar y custodiar las colecciones especiales [de la BDESSalud].
- Motivar y estimular el hábito de la lectura y la investigación, así como fomentar en la comunidad, cultura para el buen uso y aprovechamiento de los recursos bibliotecarios y proporcionar al usuario un ambiente adecuado.

Los servicios bibliotecarios que ofrece la BDESSalud son:

- Préstamo externo
- Préstamo interno o en sala
- Renovación en línea empleando el descubridor PRIMO
- Reserva/Apartado de materiales en línea empleando el descubridor PRIMO<sup>27</sup>

---

<sup>26</sup> Para conocer la totalidad de las normas de operación del SUBA, véase el Anexo 3. Normas de operación del SUBA-UACH.

<sup>27</sup> Este servicio se implementó recientemente (2019).

- Servicios electrónicos de información – Bases de datos de texto completo y referencial
- Formación de usuarios
- Catálogo en línea (OPAC)
- Consulta / Referencia
- Alerta bibliográfica o boletín de difusión de nuevas adquisiciones
- Préstamo de equipo de cómputo en sala
- Préstamo de atriles de acrílico para mesas de estudio en sala
- Préstamo de salas de audiovisuales, de estudio y cubículos
- Préstamo de auditorio (un cañón, pantalla, pódium y 70 sillas)
- Internet inalámbrico

Los usuarios de la BDESSalud son los estudiantes, docentes y personal administrativo de la UACH, aunque principalmente ofrece el servicio a los miembros de la comunidad académica de la DES Salud, más específicamente a usuarios de la Facultad de Enfermería y Nutriología (FEN) y la Facultad de Medicina y Ciencias Biomédicas (FMCB). La Tabla 7 ofrece un panorama en cuanto al tamaño de la base de usuarios de la BDESSalud.

**Tabla 7. Base de usuarios de la BDESSalud**

<b>Unidad académica</b>	<b>Número de estudiantes</b> (Nivel Técnico, Licenciatura, Maestría, Doctorado, Especialidades y Virtual)	<b>Número de docentes</b> (Tiempo completo, Medio tiempo y Hora clase)	<b>Número de administrativos</b> (Sindicalizados y De confianza)	<b>TOTAL</b>
<b>Facultad de Enfermería y Nutriología</b>	1,476	148	64	1,688
<b>Facultad de Medicina y Ciencias Biomédicas</b>	2,623	425	100	3,148
<b>TOTAL</b>	4,099	573	164	<b>4,836</b>

Fuente: UACH, 2020b, *Estadística básica* (pp. 13, 69 y 91).

Las características que tienen las instalaciones de la BDESSalud para el ofrecimiento de sus servicios son las siguientes: Su superficie total es de 3,591 m<sup>2</sup>, en la que incluye un área de estudio con un total de 561 espacios para estudiantes, de los cuales hay 132 mesas para cuatro personas cada una. Hay 10 cubículos de estudio en grupo, de los



cuales tres son para seis personas y siete para cuatro personas. Todos cuentan con pizarrones e instalaciones eléctricas para que los usuarios conecten sus equipos de cómputo. También hay un área con cubículos de estudio individual con 62 espacios. Los usuarios pueden emplear la conexión a Internet inalámbrico dentro de las instalaciones para sus dispositivos móviles. Se tienen dos laboratorios de cómputo, uno con 22 equipos y otro con 113, todos con acceso a Internet. La BDESSalud cuenta con un sistema de calefacción y refrigeración con 10 controles distribuidos en las diferentes áreas de la biblioteca. Con respecto a la iluminación hay 156 lámparas, de la cuales existen 133 lámparas fluorescentes y 10 lámparas de emergencia. En cuanto a la seguridad, se tienen 19 extintores, 10 hidrantes, detectores de humo, salida de emergencia y señalética ad hoc. El tipo de estantería es abierta<sup>28</sup> y existe una computadora de escritorio en el área de acervo para búsqueda del catálogo en línea.

Con respecto al personal de la biblioteca, la BDESSalud tiene una jefa del área de servicios bibliotecarios con el grado de Maestría en Bibliotecología y Ciencias de la Información y una jefa del área administrativa con Licenciatura en Administración de Empresas y Maestría en Recursos Humanos. Para el servicio de circulación en turno matutino se tienen cuatro bibliotecarios. Además, hay un bibliotecario en el área de préstamo de equipo de cómputo y cubículos, así como un bibliotecario para proceso físico y control de material documental que se envía o recibe de la Unidad de Procesos Técnicos. En el turno vespertino trabajan tres bibliotecarios en circulación y otro bibliotecario que está encargado de capacitar al resto del personal en las labores de reparación del material documental dañado.

En lo referente al proceso técnico del material documental, la *Historia de las bibliotecas de la Universidad Autónoma de Chihuahua* indica que [el SUBA] se maneja “centralizadamente con relación a los procesos técnicos de clasificación y catalogación, para que puedan ser consultados sus acervos desde cualquier instancia universitaria. A través de un sistema de redes se dispone de consulta al acervo mediante el [...] sistema

---

<sup>28</sup> Al 2021 debido a la pandemia COVID, la estantería fue cerrada temporalmente hasta que existiese un control de la contingencia.

Aleph<sup>29</sup>, de sus materiales documentales y bases de datos electrónicos, optimizando las posibilidades de aprovechar la información existente” (Aceves, 2002, p.3). Los materiales se encuentran clasificados según el Sistema de Clasificación Decimal Dewey (SCDD), en su versión 21 y catalogados<sup>30</sup> con base en la segunda edición revisada de las Reglas de Catalogación Angloamericanas (ACRR2)<sup>31</sup>, además se siguen los lineamientos del Formato MARC 21. De acuerdo con datos ofrecidos por los reportes internos de LogAleph, al mes de octubre de 2021 (correspondientes al material de la BDESSalud), ésta contaba con un total de 9,380 títulos y 19,396 volúmenes, de los cuales 2,852 títulos eran de reserva, 398 títulos de consulta, 122 eran tesis y el resto pertenecía a la colección general. Con respecto a los recursos digitales disponibles<sup>32</sup>, los usuarios de la BDESSalud emplean recursos de acceso abierto identificados y seleccionados por la jefatura de la biblioteca, contándose entre ellos: PubMed, Biblioteca Virtual en Salud (BVS), BVS Enfermería, Cochrane, DOAJ, DecS y el Catálogo Maestro de Guías de Prácticas Clínicas del Centro Nacional de Excelencia Tecnológica en Salud (CENETEC) de la Secretaría de Salud.

Una de las principales actividades para cumplir con las metas planteadas y ofrecer los servicios a la comunidad usuaria del SUBA es su proceso de automatización. Debe

---

<sup>29</sup> Aleph es un sistema integrado de gestión bibliotecaria (SIGB) que permite la automatización de las actividades diarias en centros de información. Las siglas significan Programa Hebreo Expandible para Automatización de Bibliotecas (*Automated Library Expandable Program Hebrewum*). Es un software de gestión de centros de información creado por la Universidad de Israel en 1986 y comercializado hoy en día por ExLibris, compañía israelí que a partir de 2015 fue adquirida por ProQuest, convirtiéndose así en su subsidiaria. ProQuest es una empresa que ofrece una amplia gama de productos y servicios de contenido para todo tipo de bibliotecas (Breeding, 2020). A partir de diciembre de 2021, ProQuest fue adquirida a su vez por la compañía Clarivate (Clarivate, 2021). En el futuro inmediato no se espera que haya modificaciones o cambios con respecto al producto Aleph, por lo que se prevé se siga contando con el soporte técnico necesario brindado por la representación de ProQuest para Latinoamérica.

<sup>30</sup> El MDoc está catalogado en el segundo nivel de descripción.

<sup>31</sup> Durante una reunión de trabajo del CCSUBA llevada a cabo en noviembre de 2021, se tomó la decisión de dejar de lado el trabajo con las ACRR2 e iniciar la catalogación del MDoc de nuevo ingreso considerando el código de catalogación *RDA* (siglas en inglés de *Resource Description and Access*). Lo anterior no incluirá un trabajo retrospectivo de la colección ya existente en las bibliotecas del SUBA.

<sup>32</sup> Antes de 2019, el SUBA-UACH formaba parte activa del Consorcio Nacional de Recursos de Información Científica y Tecnológica (CONRICyT). Por lo anterior, los usuarios de la BDESSalud tenían acceso a bases de datos especializadas como *American Chemical Society*, *Cambridge University Press*, *BioOne*, *Elsevier*, *Nature*, *Science AAAS*, *Springer*, *Thomson-Reuters*, *WILEY Subscription Services*, *EBSCO* y *Gale Cengage Learning*, entre otras. Sin embargo, luego de las restricciones al presupuesto federal asignado a CONRICyT dadas a conocer en 2018, la UACH se vio impedida a continuar participando en dicho Consorcio, por lo que el abanico de posibilidades de consulta de fuentes electrónicas para sus usuarios se restringió de manera significativa.

recordarse que al inicio de sus funciones el SUBA empleó el sistema Logicat para la automatización y puesta en práctica de sus actividades. En 2005 se realizó la migración de Logicat al sistema Aleph en su versión 14.1. En el transcurso del mismo año esta versión de Aleph fue actualizada a la versión 16.1, además de que se implementó el generador de reportes LogAleph<sup>33</sup>, a partir de lo cual se fueron configurando informes que resultaran de utilidad para el análisis de la operación del sistema. Los análisis estructurados en LogAleph han sido enfocados desde entonces principalmente a los temas de circulación y procesamiento de material. Cabe mencionar que del sistema Aleph únicamente se emplean los módulos de Catalogación, Circulación y Catálogo en línea (OPAC). Para 2021, la versión vigente del Aleph en el SUBA era la 23.1. Dentro de los servicios añadidos en los últimos años al Aleph contratados por el SUBA se encuentra el servicio de PRIMO, el cual se define como un servicio de descubrimiento o metabuscador de fuentes de información impresa y digital, sin importar su formato o ubicación.

Con respecto al financiamiento del SUBA, al momento no cuenta con un presupuesto oficial determinado por la UACH, por lo cual el principal medio para hacerse de recursos es a partir de su participación en proyectos de financiamiento federal como el Programa Integral de Fortalecimiento Institucional (PIFI), el Programa de Fortalecimiento de la Calidad en Instituciones Educativas (PROFOCIE) y el Programa de Fortalecimiento de la Calidad Educativa (PFCE), los cuales tienen como objetivo el apoyar a la ampliación y modernización de los sistemas integrales de información y la infraestructura académica de las Instituciones de Educación Superior (IES). Dentro de los requisitos planteados para acceder a estos apoyos federales, las IES deben someterse a la certificación de procesos estratégicos, entre los que se encuentran los referentes al funcionamiento de las bibliotecas. En el caso del SUBA, se ha caracterizado por ser pionero y mantener una activa participación en el Sistema de Gestión Universitario (SGU) de la UACH, el cual promueve la mejora continua para satisfacer las necesidades de la comunidad

---

<sup>33</sup> Servidor de reportes diseñado por los ingenieros Enrique Martínez Zúñiga y Federico Portas Lagar, del Instituto Tecnológico y de Estudios Superiores de Occidente (ITESO). (Martínez-Zúñiga y Portas-Lagar, 2002).

universitaria. El SGU está diseñado para dar cumplimiento a las normas y a los requisitos legales y reglamentarios aplicables señalados en la Tabla 8.

**Tabla 8. Normatividad bajo la cual opera el SUBA**

Norma	Descripción
<b>ISO 9001:2015</b>	<p>Estándar internacional de carácter certificable que regula los sistemas de gestión de la calidad. El objetivo del SUBA bajo esta norma es ofrecer una cobertura de servicios de calidad con enfoque a la comunidad universitaria, para lo cual tiene al momento certificado el <b>Proceso de Servicios Bibliotecarios</b>, del cual se desprenden los siguientes subprocesos (también certificados):</p> <ol style="list-style-type: none"> <li>1. Automatización</li> <li>2. Desarrollo de colecciones</li> <li>3. Formación de usuarios</li> <li>4. Préstamo de instalaciones</li> <li>5. Procesamiento de material documental</li> <li>6. Recursos electrónicos</li> <li>7. Circulación</li> </ol>
<b>ISO 14001:2015</b>	<p>Esta norma delimita los requisitos para establecer un sistema de gestión ambiental eficiente, lo cual otorga a las organizaciones un marco para proteger el medio ambiente y responder a las condiciones ambientales cambiantes, siempre guardando el equilibrio con las necesidades socioeconómicas.</p>
<b>ISO 45001-2018</b>	<p>Certificación de sistemas de gestión de la seguridad y salud en el trabajo. Es una norma internacional que brinda un marco para identificar, controlar y disminuir los riesgos asociados con la salud y la seguridad en el lugar de trabajo.</p>

Fuente: Información obtenida durante una entrevista con la jefa del área de servicios bibliotecarios de la BDESSalud.

Dado que el objetivo de esta tesis es incidir en la creación de un modelo de bibliominería aplicable a las actividades bibliotecarias para apoyar a la toma de decisiones, se hará énfasis en las metas e indicadores relacionados al proceso general de servicios bibliotecarios que se señalan en la Tabla 9.

**Tabla 9. Metas e indicadores del Proceso General de Servicios Bibliotecarios del SUBA**

Objetivo	Meta	Indicador	Fórmula	Periodicidad	Acciones para alcanzarlos	Responsable	
<b>Servicios Bibliotecarios</b>							
Ofrecer una cobertura de servicios de calidad con enfoque a la comunidad universitaria.	90% de satisfacción de los usuarios de los servicios bibliotecarios.	Cantidad de encuestas respondidas en el rango más alto definido en los criterios.	Cantidad de encuestas en rango *100 / Cantidad de encuestas aplicadas	Anual	Rediseñar la encuesta y la metodología de aplicación. Análisis de los resultados.	Grupo de trabajo	
	Reducir las multas.		Cantidad de multas actuales – cantidad de multas en el semestre anterior.	Semestre	Capacitar al personal, en circulación. Mantener vigente la información con los usuarios.	Coordinadora del SUBA	
	100% de solicitudes de cursos inducción y BD, atendidas.		Cantidad de solicitudes de cursos inducción y BD, atendidas.	Semestre	Contar con un grupo de Jefes de Biblioteca que puedan apoyar en los cursos de BD. Vincular los cursos de BD e Inducción en las materias transversales de manejo de la información.		
	Cantidad de servicios otorgados.		Cantidad de servicios otorgados ≥ Cantidad de servicios otorgados el año anterior.	Año	Difundir los servicios a través de todos los medios disponibles en la institución (impresos / digitales)		
	1% Máximo de Salidas no conformes.		Total de salidas no conformes * 100/ Total de servicios	Mensual	Capacitación del personal para identificar y registrar las salidas no conformes.		
	5% Máximo de error de devolución de MD		Cantidad de errores *100/ Cantidad de devoluciones	Mensual	Capacitación del personal para corregir los errores de circulación.		
	<b>Procesos técnicos y desarrollo de colecciones</b>						
	Actualizar la colección.	80% material documental actualizado en un año.		MD Adquirido = (Cantidad de material adquirido entre el rango de edición de 5 y 10 años) entre (el total de material adquirido) x 100	Semestral	Solicitar material documental por donación o compra en el rango de fecha de edición o reimpresión.	
		1% máximo de pérdida de material documental.		Total de volúmenes perdidos *100/ Total de volúmenes.	Anual	Realizar en tiempo y forma el inventario en todo el SUBA.	
	<b>Sistemas</b>						
Mantener en un 97% como mínimo, el tiempo del funcionamiento efectivo del servidor e interfaz del Sistema Aleph.	97% Tiempo de los servicios activos		Funcionamiento = (Total de tiempo de servicios activos) – (Total de tiempo de servicios inactivos)	Mensual	Monitorear diariamente los servicios del servidor (conexión y respaldo)		

Fuente: *Sistema de Gestión Universitario. Matriz de objetivos.* Universidad Autónoma de Chihuahua, 2020 (p. 2)

Considerando la normativa de operación bajo la cual funciona la BDESSalud, así como los servicios que ofrece a su comunidad y la aportación que de ella se espera para la meta global de Servicios Bibliotecarios del SUBA, se puede determinar que sus procesos fundamentales son los relacionados con el uso o circulación del material documental (MDoc) tanto impreso como electrónico, la impartición de programas de alfabetización informacional, así como el préstamo de computadoras, cubículos, instalaciones y atriles. Con lo anterior, es posible enlistarlos de la siguiente manera:

- Préstamo externo de MDoc
- Préstamo interno de MDoc
- Renovación de MDoc

- Renovación en línea de MDoc
- Devolución de MDoc
- Cobro de multas por retraso en entrega de MDoc
- Consulta de MDoc electrónico
- Programas de alfabetización informacional
- Préstamo de computadoras
- Préstamo de cubículos
- Préstamo de instalaciones

La información hasta ahora recabada resultará determinante en el proceso de diseño y estructuración del DW que apoye a la toma de decisiones de la BDESSalud, cuyos pormenores se detallan en el siguiente apartado.

## 2.2 Diseño del data warehouse

Los almacenes de datos (DW) son el núcleo de la bibliominería: permiten a la bibliometría y a la MD operar dentro del mismo espacio de datos, de modo que los bibliotecarios pueden descubrir más fácilmente patrones que abarcan tanto la creación como el uso de las obras (Nicholson, 2006d, p. 6). La metodología elegida para el desarrollo del DW es Hefesto, pues permite la construcción de DW “de forma sencilla, ordenada e intuitiva”. Tal propuesta “está fundamentada en una extensa investigación, comparación de metodologías existentes y el aporte de experiencias [...] en procesos de diseño e implementación de DW” (Bernabeu y García, 2017, p. 124).

Entre las principales características de la metodología Hefesto, se encuentran las siguientes:

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- La piedra fundamental la constituyen los requerimientos de los usuarios, por lo cual, se adapta con facilidad y rapidez a los cambios del negocio<sup>34</sup>.

---

<sup>34</sup> En este caso, los sujetos a quienes se orienta el diseño del DW y la implementación del modelo de bibliominería son las jefas (tanto de servicios bibliotecarios como administrativa) de la BDESSalud, así como a la Coordinación del SUBA-UACH.

- Reducen la resistencia al cambio.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Es independiente del software/hardware que se utilicen para su implementación (Bernabeu y García, 2017, p. 126).

Dentro de las ventajas de construir un DW para la BDESSalud se cuentan el que ayudará a transformar su información operativa en información analítica orientada a la toma de decisiones, además de que posibilita la generación de reportes dinámicos, pues actualmente sólo se cuenta con información estadística que no ofrece facilidad de análisis por su falta de integración. Se espera que lo anterior apoye en el logro de los objetivos de la BDESSalud y a su mejora continua.

A grandes rasgos, Hefesto se compone de cuatro pasos: el análisis de los requerimientos de los usuarios del DW, el análisis de las fuentes de datos, el desarrollo del modelo lógico del DW y la integración de datos. Estos pasos se detallan en los siguientes apartados.

### **2.2.1 Análisis de requerimientos**

En esta primera etapa, se realizó la labor de identificar, conocer y reunir las necesidades de información de los usuarios del DW, lo cual se llevó a cabo a partir del desarrollo de las siguientes tareas:

- **Revisión de actas de reuniones del Consejo Consultivo del SUBA**

Como se mencionó anteriormente, la BDESSalud depende directamente de la Coordinación General del SUBA, por lo cual se consideró importante conocer cuáles eran los temas de mayor relevancia para el sistema y que tienen repercusiones para las bibliotecas que lo conforman. Esto fue posible gracias al análisis de 72 actas de las reuniones llevadas a cabo por el CCSUBA, que abarcan el periodo de enero de 2005 a

septiembre de 2019. La Tabla 10 muestra los principales temas tratados durante las reuniones documentadas en dichas actas.

**Tabla 10. Principales temas y subtemas tratados en reuniones del CCSUBA**

Temas	Subtemas
<b>Evaluación</b>	<ul style="list-style-type: none"> <li>- Procedimientos del Sistema de Gestión Integral de Calidad.</li> <li>- Cambio del Sistema de Gestión Integral al Sistema de Gestión Universitario.</li> <li>- El papel de las bibliotecas en la certificación de carreras.</li> <li>- Concentración de estadísticas SEP-INEGI.</li> <li>- Encuesta de satisfacción de usuarios.</li> <li>- Estadísticas de préstamo.</li> <li>- Estadísticas del material procesado (altas, bajas y modificaciones).</li> <li>- Evaluación de bases de datos.</li> <li>- Evaluación de colecciones.</li> <li>- Indicadores de procesos: diseño, inclusión, modificaciones y exclusiones.</li> <li>- Mejora del desempeño del sistema.</li> <li>- Participación en auditorías.</li> <li>- Préstamos vencidos.</li> <li>- Procedimiento e indicador de circulación del material documental.</li> <li>- Reportes de uso del acervo.</li> </ul>
<b>Sistemas Aleph y LogAleph</b>	<ul style="list-style-type: none"> <li>- Funcionamiento y operatividad de los sistemas.</li> <li>- Actualizaciones de los sistemas.</li> <li>- Capacitación del personal en el uso apropiado de los sistemas.</li> <li>- Pérdida de información del Aleph por fallas en el servidor.</li> <li>- Confiabilidad de los datos de los reportes del LogAleph.</li> <li>- Lograr la uniformidad de la base de datos.</li> <li>- Uniformar criterios de llenado de registros de usuarios.</li> <li>- Depuración de la base de datos de usuarios.</li> <li>- Depuración de los reportes del LogAleph.</li> </ul>
<b>Material documental</b>	<ul style="list-style-type: none"> <li>- Diseño de metodología y estrategias para el desarrollo de colecciones.</li> <li>- Determinación de MDoc que apoya a cada programa académico.</li> <li>- Interacción de los jefes de biblioteca con los secretarios académicos y los docentes de las unidades académicas para evaluar y actualizar la bibliografía.</li> <li>- Descarte de materiales.</li> <li>- Edad media de la colección.</li> <li>- Inventario del material documental.</li> <li>- Material con cero usos.</li> <li>- Material rezagado de Procesos Técnicos.</li> <li>- Robo de material documental.</li> </ul>
<b>Personal bibliotecario</b>	<ul style="list-style-type: none"> <li>- Perfil de puestos para jefes de biblioteca y bibliotecarios.</li> <li>- Apuntalamiento de su formación académica.</li> </ul>



Temas	Subtemas
	<ul style="list-style-type: none"> <li>- Capacitación continua y especializada.</li> <li>- Compromiso y responsabilidad del personal con las labores del SUBA.</li> </ul>
<b>Ingresos</b>	<ul style="list-style-type: none"> <li>- Creación de cuentas y subcuentas para la identificación de ingresos de cada biblioteca por concepto de servicios bibliotecarios.</li> <li>- Revisión del procedimiento de multas.</li> <li>- Control de pagos y cancelaciones de multas.</li> <li>- Identificación de ingresos de cada biblioteca por concepto de multas.</li> <li>- Participación del SUBA en proyectos de financiamiento federal.</li> <li>- Aprobaciones de los ingresos por trámite de titulación en el rubro de donación a la biblioteca en especie.</li> </ul>
<b>Usuarios</b>	<ul style="list-style-type: none"> <li>- Cursos de inducción.</li> <li>- Cursos de formación de usuarios.</li> <li>- Capacitación en el uso de bases de datos.</li> <li>- Identificación de usuarios morosos para bloqueo.</li> </ul>
<b>Recursos electrónicos</b>	<ul style="list-style-type: none"> <li>- Capacitación del personal del SUBA en el uso de bases de datos.</li> <li>- Diseño del portal electrónico del SUBA.</li> <li>- Habilitación del repositorio digital de tesis de la UACH.</li> </ul>
<b>Reglamento</b>	<ul style="list-style-type: none"> <li>- Definición del Reglamento del SUBA y búsqueda de su aprobación por el Consejo Universitario de la UACH.</li> </ul>

---

Fuente: Elaboración propia.

- **Encuesta para profesionales de la información de bibliotecas académicas**

Se realizó una encuesta empleando la herramienta digital *QuestionPro* a 54 bibliotecarios mexicanos de diversas instituciones (incluyendo a 11 jefes de biblioteca y jefes de unidad del SUBA) para conocer cuál es el comportamiento informativo que ellos, como profesionales de la información, despliegan cotidianamente para hacer frente a las exigencias del servicio en sus centros de trabajo. El objetivo fue determinar cuáles son las necesidades de información que tienen con respecto a su trabajo y cuál es el tipo de información o datos que utilizan para satisfacerlas. Así, se compartió vía correo electrónico el enlace al cuestionario<sup>35</sup>, instrumento que constó de 23 preguntas: 7 de

---

<sup>35</sup> Véase Anexo 4. Encuesta sobre Conducta Informativa de Profesionales de la Información.

respuesta abierta, 4 de respuesta semiabierta y 12 de respuesta cerrada. La información solicitada fue con respecto a las características sociodemográficas de los participantes: grupo de edad, sexo, grado de estudios, idiomas que dominan, centro y puesto de trabajo. Asimismo, se les solicitó una breve descripción de sus actividades laborales, años de experiencia laboral, indicar su asiduidad a la consulta de publicaciones periódicas relacionadas con su trabajo, cuáles eran los temas de interés en su área laboral, las fuentes de información que consultan habitualmente, su pertenencia a organizaciones relacionadas con su trabajo, su valoración de las fuentes de información que emplean considerando elementos como familiaridad, relevancia, confiabilidad, valor, interés, accesibilidad/disponibilidad y, por último, se les solicitó señalar cuáles son los elementos que consideran le confieren utilidad a un recurso informativo. De los hallazgos realizados luego de llevar a cabo la consulta, resaltaremos aquellos que tienen relación al SUBA-UACH. Así, en cuanto a la información que requieren quienes toman las decisiones y la información que emplean para el desarrollo de sus actividades cotidianas, resaltan los siguientes temas en orden de importancia:

- Administración bibliotecaria
- Evaluación de la calidad de los servicios bibliotecarios
- Desarrollo de colecciones
- Evaluación de recursos de información digitales
- Alfabetización informacional – Formación de usuarios
- Atención a usuarios
- Mercadotecnia aplicada a bibliotecas (Difusión de recursos y servicios)
- Administración de personal
- Desarrollo y capacitación continua del personal
- Tecnología e innovación en bibliotecas

La mayor parte de la información que ocupan la obtienen del sistema de automatización de biblioteca (Aleph), reportes estadísticos obtenidos de LogAleph<sup>36</sup>, documentos del

---

<sup>36</sup> Del total de los encuestados (54 bibliotecarios), el 40% considera al SIGB como una fuente de información útil para resolver problemas y tomar decisiones en su área de trabajo, y el 36% consideran útiles las estadísticas del trabajo en bibliotecas, también para los mismos fines.

Sistema de Gestión Universitario (repositorio UNIQ), manuales de procedimiento internos, asistiendo a reuniones de trabajo, consultando a colegas, realizando búsquedas en Internet, estudiando la legislación, normativas, políticas o reglamentos en materia bibliotecológica, consultando periódicamente a los usuarios de biblioteca y realizando búsquedas sobre temas de interés en bases de datos especializadas.

Se constató que las estadísticas del trabajo en bibliotecas constituyen un recurso habitualmente disponible, con el cual están muy familiarizados y les es asignada mucha relevancia y valor, dado que para ellos ofrecen una visión muy confiable de las actividades realizadas en los centros de información. Todo ello se refleja en el amplio interés de los tomadores de decisiones en hacer uso de estas herramientas en su trabajo diario.

La **familiaridad** en cuanto al uso de fuentes de información es mayor con las búsquedas en Internet, las reuniones de trabajo y el uso de los manuales de procedimiento internos. En cuanto a **relevancia** destacaron las bases de datos especializadas, los cursos y/o talleres de capacitación y el uso de los sistemas integrales de gestión bibliotecaria (en adelante SIGB). Las fuentes de información que les brindan más **confiabilidad** resultaron ser SIGB, las bases de datos especializadas y los manuales y reglas de catalogación y clasificación, seguidas de la literatura especializada en bibliotecología (formato digital e impreso). Las fuentes con mayor **valor** resultaron ser los SIGB, las bases de datos especializadas y los cursos y/o talleres de capacitación. Las fuentes de información que les despiertan mayor **interés** en cuanto a uso son las bases de datos especializadas, los SIGB y la consulta a colegas. En cuanto a **accesibilidad/disponibilidad** de las fuentes, las de mayor presencia fueron las bases de datos especializadas, las búsquedas en Internet, la consulta de literatura especializada en bibliotecología (digital) y la consulta directa a colegas.

Esta actividad brindó perspectiva en cuanto a las necesidades de información más habituales de tomadores de decisiones en bibliotecas y las aproximaciones que éstos tienen a los recursos de información. Lo siguiente fue averiguar directamente con miembros del CCSUBA y la BDESSalud cuáles eran sus necesidades específicas, por lo cual se procedió a llevar a cabo una etapa de entrevistas.

- **Entrevistas con miembros del CCSUBA y de la BDESSalud**

Se condujeron entrevistas semiestructuradas con personal de la Unidad del Sistema de Gestión del SUBA, con encargados de la Unidad de Sistemas del SUBA, así como de la Jefatura de Servicios Bibliotecarios de la BDESSalud. Las preguntas planteadas tuvieron como objetivo la delimitación de los principales procesos del SUBA y en particular de la BDESSalud, conocer el modus operandi de los servicios ofrecidos a sus usuarios, la modalidad de registro de las transacciones derivadas de tales servicios, las principales características de los sistemas Aleph y LogAleph, así como dudas puntuales derivadas del análisis de documentos referentes al sistema. La relevancia de esta actividad radica en que permitió hacer un reconocimiento de cuáles son los principales requerimientos de información y datos de las personas que emplearán el DW.

- **Revisión de documentación referente a la BDESSalud**

Se tuvo acceso a informes de periodicidad mensual y anual que son entregados por la jefatura de la BDESSalud tanto a las autoridades académicas de la FEN y de la FMCB como a la Unidad de Concentración de Datos del SUBA y a la Coordinación General del SUBA junto con un informe escrito. Los informes están constituidos por concentrados de estadísticas generales cuya selección se derivó de los indicadores que se manejan en el Sistema de Gestión Universitario, así como datos que, para la jefatura de BDESSalud, le son relevantes para conocer el funcionamiento del centro de información. Los datos considerados para la elaboración de estos reportes son los presentados en la Tabla 11.

**Tabla 11. Datos relativos a las Estadísticas Generales de la BDESSalud**

<b>Datos</b>	<b>Proceso asociado</b>	<b>Fuente de los datos</b>
Consulta externa (préstamos externos)	Circulación del MDoc	Tablas Oracle del sistema Aleph. Reportes del LogAleph.
Consulta en sala o interna (préstamos internos)	Circulación del MDoc	Tablas Oracle del sistema Aleph. Reportes del LogAleph.
Renovaciones (hechas por el personal bibliotecario)	Circulación del MDoc	Tablas Oracle del sistema Aleph. Reportes del LogAleph.
Renovaciones en línea (hechas de manera autónoma por los usuarios)	Circulación del MDoc	Tablas Oracle del sistema Aleph. Reportes del LogAleph.
Devoluciones de MDoc	Circulación del MDoc	Tablas Oracle del sistema Aleph. Reportes del LogAleph.
Multas FEN	Circulación del MDoc	Tablas Oracle del sistema Aleph. Reportes del LogAleph.

<b>Datos</b>	<b>Proceso asociado</b>	<b>Fuente de los datos</b>
Multas FMCB	Circulación del MDoc	Tablas Oracle del sistema Aleph. Reportes del LogAleph.
Consulta de material electrónico (CDs)	Circulación del MDoc	Tablas Oracle del sistema Aleph. Reportes del LogAleph.
Títulos prestados	Circulación del MDoc	Tablas Oracle del sistema Aleph. Reportes del LogAleph.
Cursos de inducción (Número de usuarios atendidos)	Impartición de programas de alfabetización informacional	Base de datos interna en Excel
Cursos de bases de datos (Número de usuarios atendidos)	Impartición de programas de alfabetización informacional	Base de datos interna en Excel
Cursos sobre manejo de administradores de citas bibliográficas (Número de usuarios atendidos)	Impartición de programas de alfabetización informacional	Base de datos interna en Excel
Otros cursos (visitas guiadas, uso de bases de datos específicas, referencia, etc. Núm. de usuarios atendidos)	Impartición de programas de alfabetización informacional	Base de datos interna en Excel
Préstamo de computadoras (Número de solicitudes)	Préstamo de instalaciones	Base de datos interna en Excel
Préstamo de cubículos individuales (Número de solicitudes)	Préstamo de instalaciones	Base de datos interna en Excel
Préstamo de cubículos grupales (Número de solicitudes y cuántas personas los utilizaron)	Préstamo de instalaciones	LogAleph
Préstamo de auditorio (Número de solicitudes y número de personas que acudieron al auditorio)	Préstamo de instalaciones	Base de datos interna en Excel
Préstamo de vestíbulo de biblioteca (Número de solicitudes y número de personas que acudieron al vestíbulo)	Préstamo de instalaciones	Base de datos interna en Excel
Préstamo de atriles (Número de solicitudes)	Préstamo de instalaciones	Base de datos interna en Excel
Donaciones de títulos	Desarrollo de colecciones	Base de datos interna en Excel
Donaciones de ejemplares	Desarrollo de colecciones	Base de datos interna en Excel
Ejemplares de MDoc que ingresaron a circulación	Desarrollo de colecciones	LogAleph – Reporte manejado por la Jefatura de Sistemas
Reparación de MDoc	Procesos técnicos	LogAleph
Expedición de cartas de donación y no adeudo	Servicios administrativos	Base de datos interna en Excel
Duplicado de cartas de no adeudo	Servicios administrativos	Base de datos interna en Excel
Accesos al Blog La Voz del Bibliotecario	Difusión de la biblioteca	Cancelado desde abril de 2021 por orden institucional
Seguimiento a quejas	Evaluación del servicio	Reportes en Word entregados a la Coordinación General del SUBA de forma mensual
Salida no conforme (Fallas en el proceso bibliotecario)	Evaluación del servicio	Reporte del SGU
Porcentaje de salida no conforme	Evaluación del servicio	Reporte del SGU
Regreso de MDoc que no está en préstamo (sin transacción, error) (Número y porcentaje)	Errores de devolución de MDoc	Tablas Oracle del sistema Aleph. Reportes del LogAleph.
Cantidad de usuarios que pasaron por arcos de seguridad	Control de entradas y salidas	Registro interno de entradas y salidas
Entrega de libros de inglés	Otros servicios	Base de datos interna en Excel
Total de servicios		Base de datos interna en Excel

Fuente: Elaboración propia a partir de la revisión del documento interno “*Estadísticas Generales Biblioteca DES de Salud*” y entrevistas hechas a la jefa de servicios bibliotecarios de la BDESSalud.

- **Análisis del proceso de toma de decisiones en la BDESSalud y el SUBA**

Para identificar las necesidades de información para la toma de decisiones fue necesario llevar a cabo una inmersión en la manera en que opera la BDESSalud para conocer cuál era su proceso de toma de decisiones. Esto fue posible gracias a la interacción con miembros tanto del SUBA como de la misma BDESSalud, así como al análisis de documentación que refleja su normativa, procesos y operaciones. En el caso de sus normas, podemos señalar que plantean pautas que apoyan, hasta cierto punto, las tareas de evaluación y toma de decisiones en la BDESSalud. En este rubro, cabe destacar que el CCSUBA (en el cual participa la BDESSalud) sostiene reuniones periódicas (aproximadamente una al mes), de las cuales se derivan discusiones y análisis de los temas de interés para el sistema, configuración de grupos de trabajo o comisiones en torno a las problemáticas detectadas, así como toma de decisiones y celebración de acuerdos que se plasman en las Actas de CCSUBA (cuya revisión se explicó anteriormente). Tanto la jefa del área de servicios bibliotecarios, como la jefa administrativa de la BDESSalud acuden a las reuniones del CCSUBA y, considerando las decisiones tomadas ahí, desarrollan el plan de trabajo para la BDESSalud. Asimismo, la jefa del área de servicios bibliotecarios lleva a cabo una labor de acercamiento y sensibilización con las autoridades académicas de las facultades de Enfermería y Nutriología y Medicina y Ciencias Biomédicas, con el fin de acercarlas al centro de información, involucrarlas de manera más activa en su dinámica de trabajo, conocer las necesidades de sus usuarios y establecer rutas de acción para satisfacerlas a partir de los servicios de la BDESSalud.

Como apoyo a estas actividades, el sistema cuenta con la Unidad de Concentración de Datos, la cual tiene como objetivo reunir la información generada por las diversas unidades y bibliotecas producto de los servicios ofrecidos a la comunidad. Lo anterior tiene como fin el constituir insumos para la elaboración de informes y concentrados de estadísticas institucionales (UACH), estadísticas para el Consejo Nacional para Asuntos Bibliotecarios de las Instituciones de Educación Superior, A. C. (CONPAB-IES), estadísticas anuales para la Secretaría de Educación Pública (SEP) e Instituto Nacional

de Estadística y Geografía (INEGI)<sup>37</sup>, estadísticas para la Dirección de Información y Planeación de la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES), Comités Interinstitucionales para Evaluación de la Educación Superior (CIEES), agencias acreditadoras de carreras<sup>38</sup>, estadísticas para proyectos federales de financiamiento, así como información solicitada por CONRICyT<sup>39</sup>, entre otros. Además de las estadísticas anteriormente señaladas, la toma de decisiones en la BDESSalud también se apoya en el generador de reportes LogAleph y en los reportes que entrega la Unidad de Sistemas a la Coordinación sobre el uso de PRIMO, con el apoyo de la herramienta PRIMO Analytics. En este punto cabe resaltar que los reportes generados por esta vía pueden ser consultados únicamente por la Unidad de Sistemas del SUBA y al momento no se ofrece una retroalimentación directa, formal, y/o sistemática con respecto a los resultados obtenidos a los jefes de biblioteca.

- **Análisis de la información derivada de la base de datos transaccional Aleph a partir del generador de reportes LogAleph**

El SIGB empleado por la BDESSalud es Aleph. Éste se divide en módulos representativos de las principales áreas de interés en una biblioteca: módulo de Catalogación, módulo de Catálogo de Acceso Público Abierto en Web (OPAC), módulo de Adquisiciones/Seriadas y módulo de Circulación. Asimismo, incluye funcionalidades de diseminación selectiva de información (DSI) y de reserva y lectura. Hasta hace algunos años el Grupo Sistemas Lógicos, S. A. de C. V. era el comercializador oficial de Aleph en México. Actualmente lo ofrece ProQuest México y, en diciembre de 2021, ProQuest fue adquirida por la empresa Clarivate (Clarivate, 2021). La estructura del sistema está fundamentada en Oracle y ofrece soporte completo en Unicode, la administración de reportes con XML (eXtensible Markup Language) y enlaces API

---

<sup>37</sup> Conocidas como “Estadísticas SEP-INEGI”, correspondientes al formato de Estadística 912 para el nivel de Educación Superior de la SEP (<https://www.f912bibliotecas.sep.gob.mx/2020/LoginBiblioteca.aspx>).

<sup>38</sup> En el caso de la FEN y la FMCB, los organismos que han acreditado algunos de sus programas académicos son el CACEI, el COMACE, el COMAEM y el CONCAPREN (todos reconocidos por el COPAES).

<sup>39</sup> Que incluyen los resultados de la aplicación de un instrumento de evaluación de bases de datos, resultados de encuestas de satisfacción de los usuarios con los servicios ofrecidos por el CONRICyT, el cálculo de los usuarios potenciales, etc.

(*Application Programming Interface* – Interfaz de programación de aplicaciones) hacia otras aplicaciones. Aleph está “basado en estándares industriales [que permiten al sistema tener la] capacidad para compartir recursos, completa conectividad e interacción transparente con otros sistemas y bases de datos” (Grupo Sistemas Lógicos, 2017). El proveedor de Aleph ofrecía como complemento analítico el sistema *Aleph Reporting Center* (ARC), sin embargo, debido al alto costo que éste implicaba, en el SUBA se tomó la decisión de implementar en el 2005 el sistema LogAleph, el cual permite generar reportes y facilita su parametrización, agrupación y configuración. Este sistema emplea una conexión bajo un estándar de acceso a bases de datos denominado *Open Database Connectivity* (ODBC), que ayuda a extraer los datos necesarios sin afectar la integridad de la base operacional (Aleph). Los reportes se generan a partir de la escritura de consultas en lenguaje de consulta estructurada (SQL por las siglas de *Structured Query Language*). Posteriormente, se deben definir los parámetros del reporte. Una vez hecho esto, tanto la consulta como los parámetros se incluyen en la configuración de LogAleph y así el reporte queda listo.

En el caso del SUBA, el LogAleph considera las siguientes categorías para agrupar los 123 reportes existentes hasta noviembre del 2021:

1. Uso del acervo (50 reportes)
2. Material vencido (7 reportes)
3. Procesos técnicos (57 reportes)
4. Servicios bibliotecarios (5 reportes)
5. Unidad de sistemas (4 reportes)

LogAleph presenta el nombre y descripción de cada reporte, así como los parámetros a considerar dependiendo del tipo de reporte (Biblioteca, fecha inicial, fecha final, hora de inicio, hora de término, matrícula del usuario, nombre de la cuenta del bibliotecario que atendió la transacción, número de IP, etc.) y las variables que se emplean para la creación del informe, que puede ser descargado tanto en formato Excel (.xlsx) como archivo de caracteres separados por comas (.csv). En el Anexo 5 se muestra un listado de los parámetros y variables incluidos en cada categoría de reportes del LogAleph del SUBA UACH.



Aparentemente la creación de cada reporte ha respondido a las demandas de información de los integrantes del CCSUBA, quienes, dependiendo de la problemática o tema que están tratando, solicitan al encargado de la Unidad de Sistemas del SUBA la delimitación de un reporte que les ofrezca la información apropiada. Sin embargo, existen informes que aun con un nombre distinto, ofrecen los mismos resultados. Debido a esto se lleva a cabo un proceso de depuración de reportes continuo, para verificar cuáles de ellos son los realmente importantes para la toma de decisiones del sistema y sus bibliotecas.

A partir de la interacción con miembros del SUBA y de la revisión conjunta de su trabajo se confirmó que la base de datos empleada para la automatización de los procesos está diseñada para el trabajo transaccional, no para el análisis de datos. También se recabaron evidencias de estadísticas que describen al sistema y a la propia BDESSalud provenientes, sobre todo, del LogAleph. Empero, no hay claridad con respecto a cómo pueden ser explotadas o empleadas de manera más completa para brindar retroalimentación suficiente en el proceso de toma de decisiones. Inclusive la disponibilidad de las estadísticas para la consulta no siempre fue plena, pues al hacer la solicitud de éstas al personal bibliotecario no se tenían los registros completos. De acuerdo con diversos autores, se reconoce “que los sistemas tradicionales de biblioteca no son suficientes para un adecuado soporte en la toma de decisiones y [...] una eficiente gestión de la biblioteca necesita de la creación de sistemas u bases de datos especializados que integren eficientemente toda la información” (Haro et al. 2014, p.246). Debido a lo anterior y a la necesidad de contar con estructuras de datos que permitan realizar un análisis más detallado y útil de los mismos, se justifica y cobra importancia el objetivo de esta tesis: diseñar un modelo de bibliominería que aporte información útil en los procesos de evaluación y toma de decisiones en bibliotecas universitarias.

- **PRIMO *Analytics***

Una más de las fuentes de información para la toma de decisiones en la BDESSalud es el sistema PRIMO, que cuenta con un apartado para realizar análisis de los datos del sistema, denominado PRIMO *Analytics*. Entre los reportes que es posible realizar en este servicio se encuentran: las acciones emprendidas por el usuario final, tipos de dispositivos empleados en las sesiones de PRIMO, lista de términos de búsqueda más populares empleados por los usuarios finales, lista de las búsquedas con cero resultados, entre otros. Sin embargo, los resultados que se comparten derivados esta aplicación de la Coordinación a los jefes de biblioteca no dejan de ser demasiado espaciados en su temporalidad y generales en su contenido.

El análisis previo ayudó a conocer cuáles son las principales características de la BDESSalud y del sistema en que se enmarca, cuáles son los temas de interés para los tomadores de decisiones, así como los objetivos, procesos y normativas que dirigen la forma en que opera. Gracias a ello, podemos deducir que los principales procesos de la biblioteca son los relacionados con el uso o circulación del material documental (MDoc), la impartición de programas de alfabetización informacional y el préstamo de computadoras, cubículos e instalaciones. Esto se tomará en cuenta en las siguientes etapas de construcción del DW. A continuación, se ofrece la delimitación de las preguntas de negocio considerando lo expuesto hasta el momento.

#### **2.2.1.1 *Preguntas de negocio***

Luego de la revisión llevada a cabo del sistema, se corroboró que su principal objetivo es satisfacer las necesidades de información de sus usuarios y contribuir al logro académico de la UACH. Recordemos que, como indican Portas y Gómez (2014), “los indicadores deben dar luz sobre la relación biblioteca-academia, de manera que se potencialice esta relación y permita establecer estrategias comunes para incentivar el uso de los recursos de información y su aplicación en las labores cotidianas de la

academia”. Así, su principal proceso es el de circulación de MDoc, pues le permite cumplir en gran medida el objetivo planteado.

En la revisión de literatura sobre aplicaciones de Ciencia de datos en las labores bibliotecológicas fue posible detectar un gran número de metodologías que permiten abordar distintos aspectos de las actividades bibliotecarias. Un atisbo de las principales variables empleadas por dichas aplicaciones puede encontrarse en el Anexo 6. Variables de interés vs variables disponibles en la BDESSalud, el cual fue elaborado luego de la revisión de los datos existentes (más no siempre disponibles en cuanto a su posibilidad de consulta y/o uso) de la BDESSalud. Ahora, la premisa principal de esta tesis fue desarrollada considerando las necesidades de información de los tomadores de decisiones de la BDESSalud y la disponibilidad de datos de la biblioteca. De esta manera, las preguntas de negocio están directamente relacionadas a las transacciones que se llevan a cabo en el proceso de circulación del MDoc impreso:

1. ¿Cuál es el **número y/o tipo de préstamos** de un determinado **MDoc**, de determinado **autor**, de un determinado **título**, con un determinado **pie de imprenta**, de un determinado **año de MDoc**, de un determinado **tema**, de determinada **colección**, de un determinado **tipo de material**, de determinado **tipo de adquisición**, a determinado **tipo de usuario**, de determinado **sexo**, de determinada **carrera**, de determinada **unidad académica**, de determinada **biblioteca de adscripción**, por un determinado **bibliotecario** en un determinado **tiempo** en la BDESSalud?
2. ¿Cuál es el **número y/o tipo de renovaciones** de un determinado **MDoc**, de determinado **autor**, de un determinado **título**, con un determinado **pie de imprenta**, de un determinado **año de MDoc**, de un determinado **tema**, de determinada **colección**, de un determinado **tipo de material**, de determinado **tipo de adquisición**, a determinado **tipo de usuario**, de determinado **sexo**, de determinada **carrera**, de determinada **unidad académica**, de determinada **biblioteca de adscripción**, por un determinado **bibliotecario**, en un determinado **tiempo** en la BDESSalud?
3. ¿Cuál es el **número de devoluciones** de un determinado **MDoc**, de determinado **autor**, de un determinado **título**, con un determinado **pie de imprenta**, de un determinado **año de MDoc**, de un determinado **tema**, de determinada **colección**, de un determinado **tipo de material**, de determinado **tipo de adquisición**, a determinado **tipo de usuario**, de determinado **sexo**, de determinada **carrera**, de

determinada **unidad académica**, de determinada **biblioteca de adscripción**, por un determinado **bibliotecario**, en un determinado **tiempo** en la BDESSalud?

4. ¿Cuál es el **número de errores de devolución de MDoc** producidos por un determinado **bibliotecario** en un determinado **tiempo** en la BDESSalud?
5. ¿Cuál es el **número de multas** asignadas de un determinado **MDoc**, de determinado **autor**, de un determinado **título**, con un determinado **pie de imprenta**, de un determinado **año de MDoc**, de un determinado **tema**, de determinada **colección**, de un determinado **tipo de material**, de un determinado **tipo de adquisición**, a determinado **tipo de usuario**, de determinado **sexo**, de determinada **carrera**, de determinada **unidad académica**, de determinada **biblioteca de adscripción**, en un determinado **tiempo** en la BDESSalud?
6. ¿Cuál es el **número de multas cobradas** a determinado **tipo de usuario**, de determinado **sexo**, de determinada **carrera**, de determinada **unidad académica**, de determinada **biblioteca de adscripción**, por un determinado **bibliotecario** en un determinado **tiempo** en la BDESSalud?
7. ¿Cuál es el **número de multas canceladas** a determinado **tipo de usuario**, de determinado **sexo**, de determinada **carrera**, de determinada **unidad académica**, de determinada **biblioteca de adscripción**, por un determinado **bibliotecario** en un determinado **tiempo** en la BDESSalud?
8. ¿Cuál es el **número de devoluciones previas** de un determinado **MDoc**, de un determinado **autor**, de un determinado **título**, de un determinado **pie de imprenta**, de un determinado **año de MDoc**, de un determinado **tema**, de determinada **colección**, de un determinado **tipo de material**, de determinado **tipo de adquisición**, de un determinado **tipo de usuario**, de determinado **sexo**, de determinada **carrera**, de determinada **unidad académica**, de determinada **biblioteca de adscripción**, por un determinado **bibliotecario**, en un determinado **tiempo** en la BDESSalud?
9. ¿Cuál es el **número de devoluciones en tiempo** de un determinado **MDoc**, de un determinado **autor**, de un determinado **título**, de un determinado **pie de imprenta**, de un determinado **año de MDoc**, de un determinado **tema**, de determinada **colección**, de un determinado **tipo de material**, de determinado **tipo de adquisición**, de un determinado **tipo de usuario**, de determinado **sexo**, de determinada **carrera**, de determinada **unidad académica**, de determinada **biblioteca de adscripción**, por un determinado **bibliotecario**, en un determinado **tiempo** en la BDESSalud?
10. ¿Cuál es el **número de devoluciones vencidas** de un determinado **MDoc**, de un determinado **autor**, de un determinado **título**, de un determinado **pie de imprenta**, de un determinado **año de MDoc**, de un determinado **tema**, de determinada **colección**, de un determinado **tipo de material**, de determinado **tipo de adquisición**, de un determinado **tipo de usuario**, de determinado **sexo**, de

determinada **carrera**, de determinada **unidad académica**, de determinada **biblioteca de adscripción**, por un determinado **bibliotecario**, en un determinado **tiempo** en la BDESSalud?

En el diseño del DW, la dimensión de tiempo es muy relevante, ya que permitirá tener varias versiones de los datos que permitirán análisis más detallados de las transacciones que se llevan a cabo en la BDESSalud. En este paso es de suma importancia el verificar que los requerimientos señalados en las preguntas de negocio estén respaldados por una fuente de la cual sea posible obtener los datos. Dado que el sistema LogAleph de la BDESSalud permite obtener datos sobre las transacciones de préstamos, devoluciones, renovaciones y multas se decidió iniciar la estructuración del DW con la información relativa al proceso de circulación del MDoc impreso. Sin embargo, como se detallará más adelante, hubo fallos en reportes de suma importancia por su contenido informativo que imposibilitaron su descarga y uso. Por lo anterior, se recurrió al sistema transaccional Aleph para extraer información de tablas específicas empleando el SQL para reunir los datos necesarios para el DW. No fue posible tener acceso a los datos referentes a la consulta de MDoc electrónico en la BDESSalud, por lo cual no se incluyó la información de este proceso en el diseño inicial del DW. También, dado que aún los sistemas de captura de datos para los procesos de impartición de programas de alfabetización informacional, préstamo de computadoras, cubículos, atriles e instalaciones están en un grado temprano de desarrollo y debido al problema de falta de consistencia de los datos hasta ahora existentes en estos rubros, se decidió dejar tales procesos pendientes para su posterior implementación en el DW de la BDESSalud.

#### **2.2.1.2** *Indicadores y perspectivas*

Ya establecidas las preguntas de negocio se procede a descomponer sus elementos para obtener los indicadores que se utilizarán y las perspectivas de análisis (o variables) que intervendrán. Se debe tener en cuenta que los indicadores son valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: saldos, importes, promedios, cantidades, sumatorias, fórmulas, etc. En cambio, las perspectivas se

refieren a las entidades mediante las cuales se quieren examinar los indicadores, con el fin de responder a las preguntas empleadas, por ejemplo: clientes, proveedores, sucursales, países, productos, rubros, etc. Cabe destacar que el tiempo suele considerarse comúnmente como una perspectiva (Bernabeu y García, 2017, p. 182). En el caso de estudio de esta tesis, los elementos a considerar de las preguntas planteadas se muestran en la Tabla 12.

**Tabla 12. Indicadores y perspectivas de las preguntas de negocio de la BDESSalud**

Indicadores	Perspectivas
1 <b>Número de préstamos</b>	de determinado <b>tipo de préstamo</b> , de un determinado <b>MDoc</b> , de determinado <b>autor</b> , de un determinado <b>título</b> , con un determinado <b>pie de imprenta</b> , de un determinado <b>año de MDoc</b> , de un determinado <b>tema</b> , de determinada <b>colección</b> , de un determinado <b>tipo de material</b> , de determinado <b>tipo de adquisición</b> , de determinado <b>tipo de usuario</b> , de determinado <b>sexo</b> , de determinada <b>carrera</b> , de determinada <b>unidad académica</b> , de determinada <b>biblioteca de adscripción</b> , por un determinado <b>bibliotecario</b> en un determinado <b>tiempo</b> en la BDESSalud
2 <b>Número de renovaciones</b>	de determinado <b>tipo de renovación</b> , de un determinado <b>MDoc</b> , de determinado <b>autor</b> , de un determinado <b>título</b> , con un determinado <b>pie de imprenta</b> , de un determinado <b>año de MDoc</b> , de un determinado <b>tema</b> , de determinada <b>colección</b> , de un determinado <b>tipo de material</b> , de determinado <b>tipo de adquisición</b> , de determinado <b>tipo de usuario</b> , de determinado <b>sexo</b> , de determinada <b>carrera</b> , de determinada <b>unidad académica</b> , de determinada <b>biblioteca de adscripción</b> , por un determinado <b>bibliotecario</b> , en un determinado <b>tiempo</b> en la BDESSalud
3 <b>Número de devoluciones</b>	de un determinado <b>MDoc</b> , de determinado <b>autor</b> , de un determinado <b>título</b> , con un determinado <b>pie de imprenta</b> , de un determinado <b>año de MDoc</b> , de un determinado <b>tema</b> , de determinada <b>colección</b> , de un determinado <b>tipo de material</b> , de determinado <b>tipo de adquisición</b> , de determinado <b>tipo de usuario</b> , de determinado <b>sexo</b> , de determinada <b>carrera</b> , de determinada <b>unidad académica</b> , de determinada <b>biblioteca de adscripción</b> , por un determinado <b>bibliotecario</b> en un determinado <b>tiempo</b> en la BDESSalud
4 <b>Número de errores de devolución de MDoc</b>	producidos por un determinado <b>bibliotecario</b> en un determinado <b>tiempo</b> en la BDESSalud
5 <b>Número de multas</b>	asignadas de un determinado <b>MDoc</b> , de determinado <b>autor</b> , de un determinado <b>título</b> , con un determinado <b>pie de imprenta</b> , de un determinado <b>año de MDoc</b> , de un determinado <b>tema</b> , de determinada <b>colección</b> , de un determinado <b>tipo de material</b> , de un determinado <b>tipo de adquisición</b> , de determinado <b>tipo de usuario</b> , de determinado <b>sexo</b> , de determinada <b>carrera</b> , de determinada <b>unidad académica</b> , de determinada <b>biblioteca de adscripción</b> , en un determinado <b>tiempo</b> en la BDESSalud

Indicadores	Perspectivas
<b>6 Número de multas cobradas</b>	a determinado <b>tipo de usuario</b> , de determinado <b>sexo</b> , de determinada <b>carrera</b> , de determinada <b>unidad académica</b> , de determinada <b>biblioteca de adscripción</b> , por un determinado <b>bibliotecario</b> en un determinado <b>tiempo</b> en la BDESSalud
<b>7 Número de multas canceladas</b>	a determinado <b>tipo de usuario</b> , de determinado <b>sexo</b> , de determinada <b>carrera</b> , de determinada <b>unidad académica</b> , de determinada <b>biblioteca de adscripción</b> , por un determinado <b>bibliotecario</b> en un determinado <b>tiempo</b> en la BDESSalud
<b>8 Número de devoluciones previas</b>	de un determinado <b>MDoc</b> , de determinado <b>autor</b> , de un determinado <b>título</b> , con un determinado <b>pie de imprenta</b> , de un determinado <b>año de MDoc</b> , de un determinado <b>tema</b> , de determinada <b>colección</b> , de un determinado <b>tipo de material</b> , de determinado <b>tipo de adquisición</b> , de determinado <b>tipo de usuario</b> , de determinado <b>sexo</b> , de determinada <b>carrera</b> , de determinada <b>unidad académica</b> , de determinada <b>biblioteca de adscripción</b> , por un determinado <b>bibliotecario</b> en un determinado <b>tiempo</b> en la BDESSalud
<b>9 Número de devoluciones en tiempo</b>	de un determinado <b>MDoc</b> , de determinado <b>autor</b> , de un determinado <b>título</b> , con un determinado <b>pie de imprenta</b> , de un determinado <b>año de MDoc</b> , de un determinado <b>tema</b> , de determinada <b>colección</b> , de un determinado <b>tipo de material</b> , de determinado <b>tipo de adquisición</b> , de determinado <b>tipo de usuario</b> , de determinado <b>sexo</b> , de determinada <b>carrera</b> , de determinada <b>unidad académica</b> , de determinada <b>biblioteca de adscripción</b> , por un determinado <b>bibliotecario</b> en un determinado <b>tiempo</b> en la BDESSalud
<b>10 Número de devoluciones vencidas</b>	de un determinado <b>MDoc</b> , de determinado <b>autor</b> , de un determinado <b>título</b> , con un determinado <b>pie de imprenta</b> , de un determinado <b>año de MDoc</b> , de un determinado <b>tema</b> , de determinada <b>colección</b> , de un determinado <b>tipo de material</b> , de determinado <b>tipo de adquisición</b> , de determinado <b>tipo de usuario</b> , de determinado <b>sexo</b> , de determinada <b>carrera</b> , de determinada <b>unidad académica</b> , de determinada <b>biblioteca de adscripción</b> , por un determinado <b>bibliotecario</b> en un determinado <b>tiempo</b> en la BDESSalud

Fuente: Elaboración propia.

Como se indicó anteriormente, tales indicadores y perspectivas pertenecen al **proceso de circulación** del MDoc impreso de la BDESSalud. De este modo, los **indicadores** identificados son:

1. Número de préstamos
2. Número de renovaciones
3. Número de devoluciones
4. Número de errores de devolución de MDoc
5. Número de multas
6. Número de multas cobradas

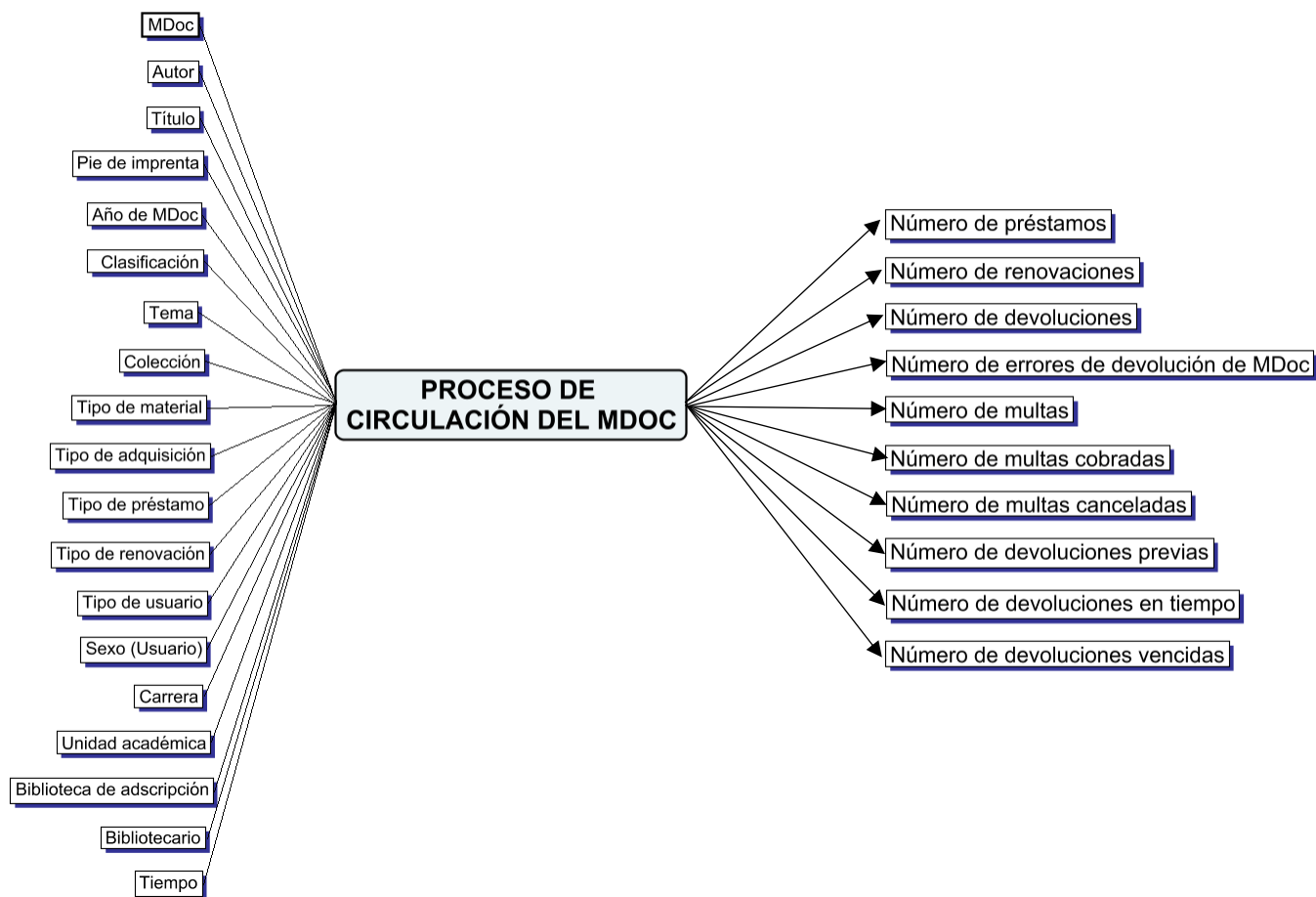
7. Número de multas canceladas
8. Tipo de devolución del MDoc

Y las **perspectivas** o **variables de análisis** a tomar en cuenta son:

1. MDoc
2. Autor
3. Título
4. Pie de imprenta
5. Año de MDoc
6. Clasificación
7. Tema
8. Colección
9. Tipo de material
10. Tipo de adquisición
11. Tipo de préstamo
12. Tipo de renovación
13. Tipo de usuario
14. Sexo (usuario)
15. Carrera
16. Unidad académica
17. Biblioteca de adscripción
18. Bibliotecario
19. Tiempo

Considerando estos elementos se delimitó un primer acercamiento visual a las variables e indicadores que se incluyeron en el modelo conceptual del DW, los cuales pueden ser apreciados en la Figura 7.





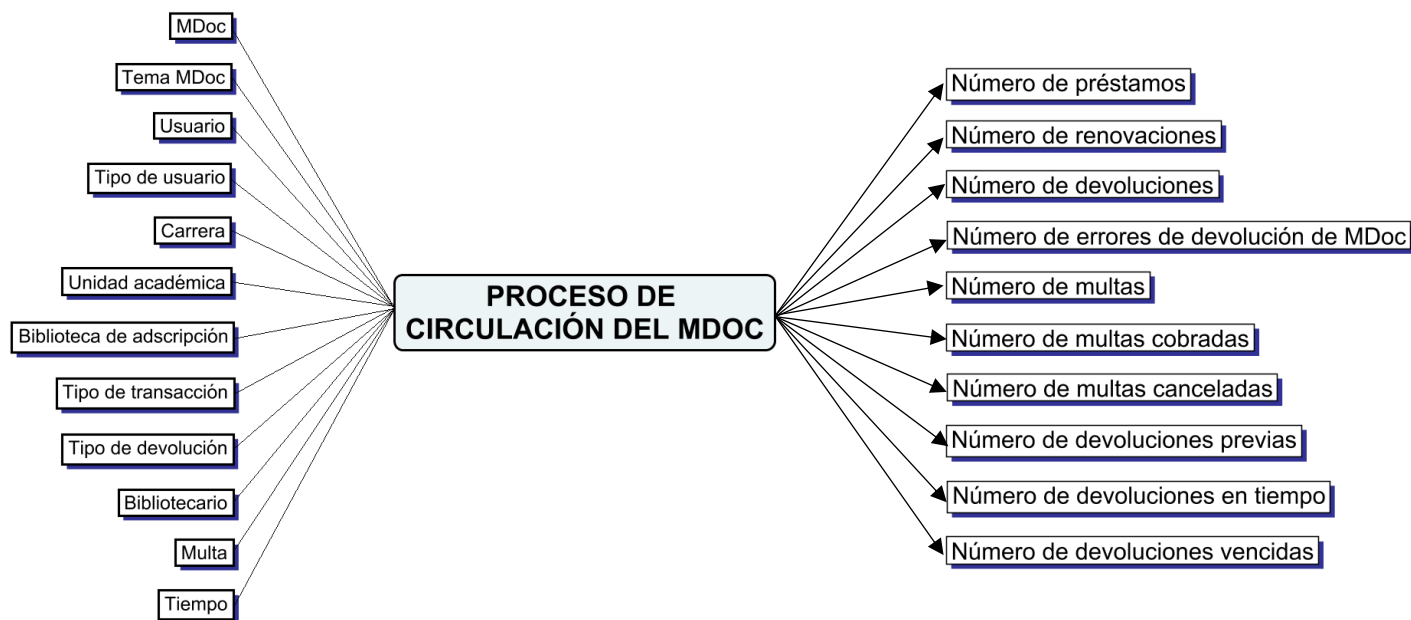
**Figura 7. Variables e indicadores del proceso de circulación de material documental**

Fuente: Elaboración propia.

### **2.2.1.3 Modelo conceptual del DW**

En este paso se considerarán los indicadores y perspectivas o variables delimitadas en el paso anterior para construir el modelo conceptual, que es una descripción de alto nivel de la estructura de la base de datos, en la cual la información es representada a partir de objetos, relaciones y atributos. El modelo conceptual señala las perspectivas de análisis (o variables) a considerar para la delimitación de los indicadores y “permite de un solo vistazo y sin poseer [...] conocimientos previos comprender cuáles serán los resultados que se obtendrán, cuáles serán las variables que se utilizarán para analizarlos y cuál es la relación que existe entre ellos” (Bernabeau y García, 2017, p. 134). Con el

fin de adaptar las variables seleccionadas para un mejor acomodo del DW, éstas fueron agrupadas en doce tablas que posteriormente fueron delimitadas como dimensiones y subdimensiones. Así, el modelo conceptual referente al proceso de circulación de MDoc de la BDESSalud se muestra en la Figura 8.



**Figura 8. Modelo conceptual del proceso de circulación de material documental**

Fuente: Elaboración propia considerando las recomendaciones de Bernabeu y García (2017).

### 2.2.2 Análisis de las fuentes de datos

En la metodología Hefesto, la siguiente etapa se orienta a la revisión de las fuentes de datos con el fin de definir su correspondencia con los requerimientos develados en la primera fase. Así es posible determinar cómo serán calculados y por ende construidos los indicadores, además de establecer el mapeo entre el modelo conceptual creado en el paso anterior y los datos de la biblioteca, lo cual permitirá ampliarlo para incidir en un buen diseño del DW.

Las fuentes de datos empleadas por bibliotecas universitarias consisten principalmente en bases de datos relacionales (sistemas OLTP), bases de datos documentales (repositorios digitales), así como sistemas de información utilizados por las instituciones que las albergan. Todas estas fuentes se relacionan entre sí para satisfacer los requerimientos operativos e informativos de la organización.

La BDESSalud cuenta con el sistema Aleph para desarrollar la mayor parte de sus procesos, además del LogAleph, que es un sistema generador de reportes. Empero, su acceso a fuentes de datos universitarias (provenientes de la UACH) es sumamente limitado y el acceso a fuentes de datos externas para la toma de decisiones, nulo. La Tabla 13 muestra el resumen de las fuentes de datos de la biblioteca bajo estudio.

**Tabla 13. Resumen de las fuentes de datos de la BDESSalud**

Tipo de fuentes de datos	Aplicación	Módulos	Base de datos	Tipo de base de datos	Principales atributos	Acceso
<b>Interna (SUBA)</b>	Aleph	Catalogación	Oracle	Documental	MDoc, autor, título, editorial, año, clasificación, etc.	Sin acceso
<b>Interna (SUBA)</b>	Aleph	Circulación	Oracle	Relacional	Usuarios (ID matrícula, nombre, apellido, carrera, etc.) fecha transacción, MDoc, préstamos, devoluciones, renovaciones, multas, cubículos, fecha préstamo, hora entrada, hora salida, ID cubículo, número de personas, bibliotecario	Acceso limitado
<b>Interna (SUBA)</b>	Aleph	Catálogo de Acceso Público Abierto en Web (OPAC)	Oracle	Relacional	<u>-Fue reemplazado por PRIMO-</u>	Acceso muy limitado
<b>Interna (SUBA)</b>	Aleph	Adquisiciones	Oracle	Relacional	<u>-El módulo no es empleado por el SUBA-</u>	Sin acceso

Tipo de fuentes de datos	Aplicación	Módulos	Base de datos	Tipo de base de datos	Principales atributos	Acceso
<b>Interna (SUBA)</b>	Aleph	Seriadas	Oracle	Relacional	<i>-El módulo no es empleado por el SUBA-</i>	Sin acceso
<b>Interna (SUBA)</b>	LogAleph	Generador de reportes sobre: uso del acervo, material vencido, procesos técnicos, servicios bibliotecarios y unidad de sistemas	MySQL	Relacional	Usuario, ID usuario, carrera, matrícula, género, biblioteca, bibliotecario, MDoc, título, autor, editorial, colección, tipo de acervo, fecha préstamo, fecha vencimiento, transacción, etc.	Acceso permitido (limitado únicamente por errores en reportes específicos)
<b>Interna (SUBA)</b>	PRIMO Analytics	Sistema de reportes de PRIMO	Oracle	Relacional	Acciones emprendidas por el usuario, tipos de dispositivos, términos de búsqueda más populares, búsquedas con cero resultados, etc.	Sin acceso
<b>Interna (SUBA)</b>	Base de datos de préstamo de equipo de cómputo	Registro de uso	Excel	Hoja de cálculo	ID usuario, matrícula, nombre, carrera, fecha, hora entrada, hora salida, bibliotecario	Acceso limitado (interno)
<b>Interna (SUBA)</b>	Encuestas de satisfacción de usuarios	Pilotaje de instrumento en FEN para aplicarse a nivel SUBA	Google forms	Hoja de cálculo	Información general del usuario, valoración del acceso a servicios bibliotecarios, calificación de los servicios bibliotecarios, grado de satisfacción con los servicios, etc.	Acceso limitado
<b>Universitaria (UACH)</b>	Sistema Estratégico de Gestión Académica (SEGA)	Inscripción	ND	Relacional	ID usuario, nombre, matrícula, género, carrera, unidad académica de adscripción, etc.	Sin acceso
<b>Universitaria (UACH)</b>	Información sobre recursos humanos de la UACH (RRHH)	Recursos humanos (académicos, administrativos y de apoyo)	Excel	Hoja de cálculo	Nombre, número de empleado, año de ingreso, vigencia, unidad académica de adscripción, etc.	Sin acceso
<b>Universitaria (UACH)</b>	UNIQ	No especificados	No especificado	Documental	Información de tipo textual. Repositorio que alberga los documentos relativos al Sistema de	Sin acceso

<b>Tipo de fuentes de datos</b>	<b>Aplicación</b>	<b>Módulos</b>	<b>Base de datos</b>	<b>Tipo de base de datos</b>	<b>Principales atributos</b>	<b>Acceso</b>
<b>Universitaria (UACH)</b>	Repositorio digital SUBA	Repositorio institucional de tesis – Proyecto Coordinado por la Dirección de Investigación y Posgrado UACH	MySQL	Relacional	Gestión Universitario de la UACH. Documentos, título, autor, resumen, fecha, palabras clave, clasificación temática, tipo de documento, unidad académica, editor, estado, formato, etc.	Sin acceso
<b>Externa</b>	Reportes de uso de las Bases de Datos del CONRICyT	Estadísticas de uso proporcionadas por el consorcio	Excel	Hoja de cálculo	Instituciones, recursos de información, descargas de documentos a texto completo, descarga de capítulos, consultas a bases de datos, herramientas clínicas, herramientas de estudio, agregadores, etc.	Sin acceso a datos de uso, que están resguardados por el CONRICyT

Fuente: Elaboración propia.

### 2.2.2.1 Hechos e Indicadores

Considerando la metodología Hefesto, continuaremos con la configuración de los hechos e indicadores del modelo de DW:

#### Conformación de indicadores

##### 1. Número de préstamos:

<b>Hecho:</b>	numero_prestamos <sup>40</sup>
<b>Función de agregación:</b>	COUNT
<b>Tipo de datos:</b>	Numéricos
<b>Fuente de datos:</b>	Tablas Oracle de Aleph
<b>Aclaración:</b>	El indicador número de préstamos representa el conteo del número de veces que un ítem (MDoc) específico ha sido prestado.

##### 2. Número de renovaciones:

<b>Hecho:</b>	numero_renovaciones
<b>Función de agregación:</b>	COUNT
<b>Tipo de datos:</b>	Numéricos
<b>Fuente de datos:</b>	Tablas Oracle de Aleph
<b>Aclaración:</b>	El indicador número de renovaciones representa el conteo del número de veces que un ítem (MDoc) específico ha sido renovado.

##### 3. Número de devoluciones:

<b>Hecho:</b>	numero_devoluciones
<b>Función de agregación:</b>	COUNT
<b>Tipo de datos:</b>	Numéricos
<b>Fuente de datos:</b>	Tablas Oracle de Aleph
<b>Aclaración:</b>	El indicador número de devoluciones representa el conteo del número de veces que un ítem (MDoc) específico ha sido devuelto.

---

<sup>40</sup> En el apartado “Hecho” de cada indicador se muestra su nombre en minúsculas, sin acentos y con sus elementos separados por guiones bajos. Esta es la manera en que los indicadores se nombran en el sistema *MySQL Workbench*, que fue empleado en la estructuración del DW de la BDESSalud. Lo anterior corresponde a un convencionalismo de ingeniería en sistemas que facilita la programación del DW y el manejo de los elementos considerados en éste.

#### 4. Número de errores de devolución de MDoc:

<b>Hecho:</b>	numero_errores_devolucion_mdoc
<b>Función de agregación:</b>	COUNT
<b>Tipo de datos:</b>	Numéricos
<b>Fuente de datos:</b>	Tablas Oracle de Aleph
<b>Aclaración:</b>	El indicador número de errores de devolución de MDoc representa el conteo del número de veces que un ítem (MDoc) específico ha sido devuelto sin estar en préstamo.

#### 5. Número de multas:

<b>Hecho:</b>	numero_multas
<b>Función de agregación:</b>	COUNT
<b>Tipo de datos:</b>	Numéricos
<b>Fuente de datos:</b>	Tablas Oracle de Aleph
<b>Aclaración:</b>	El indicador número de multas representa el conteo del número de veces que un ítem (MDoc) específico ha sido prestado y ha sido devuelto después de la fecha especificada para la devolución, por lo cual se genera una transacción de multa.

#### 6. Número de multas cobradas:

<b>Hecho:</b>	numero_multas_cobradas
<b>Función de agregación:</b>	COUNT
<b>Tipo de datos:</b>	Numéricos
<b>Fuente de datos:</b>	Tablas Oracle de Aleph
<b>Aclaración:</b>	El indicador número de multas cobradas representa el conteo del número de veces que un ítem (MDoc) específico ha generado multas y éstas han sido cobradas.

#### 7. Número de multas canceladas:

<b>Hecho:</b>	numero_multas_canceladas
<b>Función de agregación:</b>	COUNT
<b>Tipo de datos:</b>	Numéricos
<b>Fuente de datos:</b>	Tablas Oracle de Aleph

<b>Aclaración:</b>	El indicador número de multas canceladas representa el conteo del número de veces que un ítem (MDoc) específico ha generado multas y éstas han sido canceladas.
--------------------	---

### 8. Número de devoluciones previas:

<b>Hechos:</b>	numero_devoluciones_previas
<b>Función de agregación:</b>	COUNT
<b>Tipo de datos:</b>	Numéricos
<b>Fuente de datos:</b>	Tablas Oracle de Aleph
<b>Aclaración:</b>	El indicador número de devoluciones previas representa el conteo del número de veces que un ítem (MDoc) específico ha sido devuelto antes de la fecha de vencimiento establecida por el sistema.

### 9. Número de devoluciones en tiempo:

<b>Hechos:</b>	numero_devoluciones_en_tiempo
<b>Función de agregación:</b>	COUNT
<b>Tipo de datos:</b>	Numéricos
<b>Fuente de datos:</b>	Tablas Oracle de Aleph
<b>Aclaración:</b>	El indicador número de devoluciones en tiempo representa el conteo del número de veces que un ítem (MDoc) específico ha sido devuelto justo en la fecha de vencimiento establecida por el sistema.

### 10. Número de devoluciones vencidas:

<b>Hechos:</b>	numero_devoluciones_vencidas
<b>Función de agregación:</b>	COUNT
<b>Tipo de datos:</b>	Numéricos
<b>Fuente de datos:</b>	Tablas Oracle de Aleph
<b>Aclaración:</b>	El indicador número de devoluciones vencidas representa el conteo del número de veces que un ítem (MDoc) específico ha sido devuelto luego de la fecha de vencimiento establecida por el sistema.



#### 2.2.2.2 Mapeo

En la metodología Hefesto la fase de mapeo de datos implica la revisión de las fuentes de información para identificar sus características, de manera que se tenga la seguridad de contar con los datos requeridos. Este paso es necesario para establecer las vías a partir de las cuales serán obtenidos los elementos previamente definidos en el modelo conceptual.

A continuación, se establecen las correspondencias entre las tablas Oracle de Aleph (OLTP) involucradas en el proceso de circulación de material con el fin de mapear su relación con los indicadores y con las perspectivas del modelo conceptual del DW. En la documentación oficial proporcionada por ExLibris en su página electrónica fue posible encontrar las tablas de Entidad Relación de la base de datos transaccional Aleph. Sin embargo, dada la complejidad de las tablas Oracle de Aleph se consideró más apropiado el delimitar las relaciones entre las fuentes de información y las perspectivas del modelo conceptual del DW de manera tabular. Asimismo, se indica cuáles fueron los archivos independientes a la base de Aleph que fueron utilizados para la definición de las perspectivas. Es pertinente aclarar que en este mapeo se muestra el nombre otorgado a las perspectivas del modelo lógico, tanto como son manejados por los encargados de sistemas (*back end*) como los emplean los usuarios finales (*front end*). En el caso de los términos del *back end* se muestra su nombre en minúsculas, sin acentos y con sus elementos separados por guiones bajos. Esta es la manera en que estos elementos se nombran en el sistema *MySQL Workbench*, empleado en la estructuración del DW de la BDESSalud. Lo anterior corresponde a un convencionalismo de ingeniería en sistemas que facilita la programación del DW. En las próximas páginas se presentan las tablas de las distintas perspectivas e indicadores del modelo conceptual y se indica cuál fue la fuente a partir de la cual se obtuvieron los datos necesarios para la configuración del DW.

**Tabla 14. Mapeo de las fuentes de datos de las perspectivas del modelo lógico pertenecientes a la dimensión MDOC**

Dimensión: MDoc						
Fuentes de los datos				Perspectiva del modelo lógico		Ejemplo de los datos
Tabla Oracle Aleph	Columna tabla Oracle Aleph	Archivo modificado	Ubicación dentro del archivo modificado	Terminología DW		
				Back end (Encargados de sistemas) Términos para el DW	Front end (Usuarios finales)	
Z30H (Item - History)	Z30H_BARCODE			key_mdoc	ID MDoc	DS018186
Z13 (Short bibliographic record)	Z13_CALL_NO			clasificacion	Clasificación	610.736 8 PER 1998
Z13 (Short bibliographic record)	Z13_TITLE			titulo	Título	Manual de procedimientos de enfermería psiquiátrica Artemio Pérez Sosa
Z13 (Short bibliographic record)	Z13_AUTHOR			autor	Autor	Pérez Sosa, Artemio
Z13 (Short bibliographic record)	Z13_IMPRINT			pie_imprensa	Pie de imprenta	México Editorial Trillas 1998
Z13 (Short bibliographic record)	Z13_IMPRINT, SUBSTR(Z13_IMPRINT,-4,4)=AÑO			anio_mdoc	Año MDoc	1998
		ClasificacionDesagregadoTemas.xlsx	COLUMNAS: CLASE (SCDD), DIVISION (SCDD) y SECCION (SCDD)	key_tema	Tema	Technology
Z30H (Item - History)	Z30H_COLLECTION	CatCodigosOracleAleph.docx	APARTADO: TIPO DE COLECCIÓN, Desglose de los códigos alfanuméricos para la columna Z30H_COLLECTION. Válido también para la columna Z30_COLLECTION.	tipo_coleccion	Colección	Reserva
Z30H (Item - History)	Z30H_MATERIAL	CatCodigosOracleAleph.docx	APARTADO: TIPO DE MATERIAL, Desglose de los códigos alfanuméricos para la columna Z30H_MATERIAL. Coincide con la Z30_MATERIAL y la Z35_MATERIAL.	tipo_material	Tipo de material	Libro
Z30H (Item - History)	Z30H_ITEM_STATISTIC	CatCodigosOracleAleph.docx	APARTADO: TIPO DE ADQUISICIÓN, Desglose de los códigos numéricos para la columna Z30H_ITEM_STATISTIC. Es equivalente a la columna Z30_ITEM_STATISTIC.	tipo_adquisicion	Tipo de adquisición	Donación
Z30H (Item - History)	Z30H_OPEN_DATE			fecha_ingreso_mdoc	Fecha de ingreso a sistema (MDoc)	20100320
Z30H (Item - History)	Z30H_DATE_LAST_RETURN			fecha_ultima_devolucion_mdoc	Fecha de última devolución (MDoc)	20190815
Z30H (Item - History)	Z30H_ITEM_STATUS	CatCodigosOracleAleph.docx	APARTADO: ESTADO DEL MDOC, Empleado para definir el tipo de préstamo. Válidos para las columnas Z30H_ITEM_STATUS, Z30_ITEM_STATUS y Z35_ITEM_STATUS. Desglose de los códigos numéricos para la columna Z30H_ITEM_STATUS	estado_mdoc	Estado del Mdoc	3 días externo
Z30H (Item - History)	Z30H-H-REASON	CatCodigosOracleAleph.docx	APARTADO: CAMBIOS HECHOS AL MDOC, Desglose de las opciones alfanuméricas para la columna Z30H_H_REASON	mdoc_cambios	Cambios hechos al MDoc	Eliminado
Z30H (Item - History)	Z30H-H-REASON-TYPE	CatCodigosOracleAleph.docx	APARTADO: ÚLTIMO CAMBIO HECHO AL MDOC, Desglose de los códigos alfanuméricos para la columna Z30H_H_REASON_TYPE	mdoc_ultimo_cambio	Último cambio hecho al MDoc	Delete
Z30H (Item - History)	Z30H-H-DATE			fecha_baja_mdoc	Fecha de baja (MDoc)	20211027
Z30H (Item - History)	Z30H-H-HOUR			hora_baja_mdoc	Hora de baja (MDoc)	1635

Fuente: Elaboración propia.

**Tabla 15. Mapeo de las fuentes de datos de la perspectiva del modelo lógico perteneciente a la subdimensión TEMA\_MDOC**

Subdimensión: Tema_MDoc						
Fuentes de los datos				Perspectiva del modelo lógico		Ejemplo de los datos
				Terminología DW		
Tabla(s) Oracle Aleph	Columna tabla Oracle Aleph	Archivo modificado	Ubicación dentro del archivo modificado	Back end (Encargados de sistemas)	Front end (Usuarios finales)	
Z13 (Short bibliographic record), Z30 (Item), Z30H (Item - History)	Z13_CALL_NO, Z30H_CALL_NO, Z30H_CALL_NO			key_tema	Clasificación	610.736 8 PER 1998
		ClasificacionDesagregadoTemas.xlsx	COLUMNA: CLASE (SCDD)	clase_principal	Tema - CLASE (SCDD)	Technology
		ClasificacionDesagregadoTemas.xlsx	COLUMNA: DIVISION (SCDD)	division_subclase	Tema - DIVISIÓN (SCDD)	Medicine & health
		ClasificacionDesagregadoTemas.xlsx	COLUMNA: SECCION (SCDD)	seccion_grupo	Tema - SECCIÓN (SCDD)	Medicine & health

Fuente: Elaboración propia.

**Tabla 16. Mapeo de las fuentes de datos de las perspectivas del modelo lógico pertenecientes a la dimensión USUARIO**

Dimensión: Usuario						
Fuentes de los datos				Perspectiva del modelo lógico		Ejemplo de los datos
				Terminología DW		
Tabla Oracle Aleph	Columna tabla Oracle Aleph	Archivo modificado	Ubicación dentro del archivo modificado	Back end (Encargados de sistemas)	Front end (Usuarios finales)	
Z303 (Global patron information)	SUBSTR(Z303_REC_KEY,1,12)			key_usuario	ID Usuario	238589
Z303 (Global patron information)	Z303_FIELD_3			key_carrera	Clave de carrera	530001
Z303 (Global patron information)	Z303_GENDER			sexo_usuario	Sexo del usuario	Mujer
		TablaAsignarUnidadAcademica.xlsx	COLUMNA: ClaveCarreraxUsuario	key_unidad_academica	ID Unidad académica	FACULTAD DE ENFERMERIA Y NUTRIOLOGIA
Z303 (Global patron information)	Z303_HOME_LIBRARY			key_biblioteca_adscripcion	Biblioteca de adscripción	DES de Salud
Z303 (Global patron information)	Z305_BOR_STATUS			key_tipo_usuario	Tipo de usuario	Alumno
Z303 (Global patron information)	Z303_OPEN_DATE			fecha_creacion_reg_usuario	Fecha de creación del registro de usuario	20160320
Z303 (Global patron information)	Z303_UPDATE_DATE			fecha_actualizacion_reg_usuario	Fecha de actualización del registro de usuario	20180815
Z305 (Patron's local information)	Z305_EXPIRY_DATE			fecha_vencimiento_reg_usuario	Fecha de vencimiento del registro de usuario	20120630

Fuente: Elaboración propia.

**Tabla 17. Mapeo de las fuentes de datos de la perspectiva del modelo lógico perteneciente a la subdimensión TIPO\_USUARIO**

<b>Subdimensión: Tipo_Usuario</b>						
<b>Fuentes de los datos</b>				<b>Perspectiva del modelo lógico</b>		<b>Ejemplo de los datos</b>
				<b>Terminología DW</b>		
<b>Tabla Oracle Aleph</b>	<b>Columna tabla Oracle Aleph</b>	<b>Archivo modificado</b>	<b>Ubicación dentro del archivo modificado</b>	<b>Back end (Encargados de sistemas)</b>	<b>Front end (Usuarios finales)</b>	
Z303 (Global patron information)	Z305_BOR_STATUS			key_tipo_usuario	ID Tipo de usuario	01
		CatCodigosOracleAleph.docx	APARTADO: TIPO DE USUARIO, Desglose de los códigos numéricos para la columna Z305_BOR_STATUS	tipo_usuario	Tipo de usuario	Alumno

Fuente: Elaboración propia.

**Tabla 18. Mapeo de las fuentes de datos de la perspectiva del modelo lógico perteneciente a la subdimensión CARRERA**

Subdimensión: Carrera						
Fuentes de los datos				Perspectiva del modelo lógico		Ejemplo de los datos
				Terminología DW		
Tabla Oracle Aleph	Columna tabla Oracle Aleph	Archivo modificado	Ubicación dentro del archivo modificado	Back end (Encargados de sistemas)	Front end (Usuarios finales)	
Z303 (Global patron information)	Z303_FIELD_3	TablaAsignarUnidadAcademica.xlsx	COLUMNA: ClaveCarreraxUsuario	key_carrera	Clave de carrera	530001
Z303 (Global patron information)	Z303_FIELD_2	TablaAsignarUnidadAcademica.xlsx	COLUMNA: DescProgramaEducativo	<b>carrera</b>	<b>Carrera</b>	Licenciado en Enfermería
		TablaAsignarUnidadAcademica.xlsx	COLUMNAS: ClaveCarreraxUsuario, DescNivelAcademico	nivel_academico	Nivel académico	Licenciatura
		TablaAsignarUnidadAcademica.xlsx	COLUMNAS: ClaveCarreraxUsuario, DescProgramaEducativo	programa_educativo	Programa educativo	Licenciado en Enfermería

Fuente: Elaboración propia.

**Tabla 19. Mapeo de las fuentes de datos de la perspectiva del modelo lógico perteneciente a la subdimensión UNIDAD\_ACADEMICA**

<b>Subdimensión: Unidad_Academica</b>						
<b>Fuentes de los datos</b>				<b>Perspectiva del modelo lógico</b>		<b>Ejemplo de los datos</b>
				<b>Terminología DW</b>		
<b>Tabla Oracle Aleph</b>	<b>Columna tabla Oracle Aleph</b>	<b>Archivo modificado</b>	<b>Ubicación dentro del archivo modificado</b>	<b>Back end (Encargados de sistemas)</b>	<b>Front end (Usuarios finales)</b>	
		TablaAsignarUnidadAcademica.xlsx	COLUMNA: ClaveCarreraxUsuario	key_unidad_academica	ID Unidad Académica	5300
		TablaAsignarUnidadAcademica.xlsx	COLUMNAS: ClaveCarreraxUsuario, Unidad Academica	unidad_academica	Unidad Académica	Facultad de Enfermería y Nutrición

Fuente: Elaboración propia.

**Tabla 20. Mapeo de las fuentes de datos de la perspectiva del modelo lógico perteneciente a la subdimensión BIBLIOTECA\_ADSCRIPCION**

Subdimensión: Biblioteca_Adscricion						
Fuentes de los datos				Perspectiva del modelo lógico		Ejemplo de los datos
				Terminología DW		
Tabla Oracle Aleph	Columna tabla Oracle Aleph	Archivo modificado	Ubicación dentro del archivo modificado	Back end (Encargados de sistemas)	Front end (Usuarios finales)	
Z303 (Global patron information)	Z303_HOME_LIBRARY			key_biblioteca_adscricion	ID Biblioteca de adscrición	DS
		CatCodigosOracleAleph.docx	APARTADO: BIBLIOTECA DE ADSCRIPCIÓN, Desglose de los códigos alfanuméricos para la columna Z303_HOME_LIBRARY	<b>biblioteca_adscricion</b>	<b>Biblioteca de adscrición</b>	DES de Salud

Fuente: Elaboración propia.



**Tabla 21. Mapeo de las fuentes de datos de la perspectiva del modelo lógico perteneciente a la dimensión TIPO\_TRANSACCION**

Dimensión: Tipo_Transaccion						
Fuentes de los datos				Perspectiva del modelo lógico		Ejemplo de los datos
				Terminología DW		
Tabla Oracle Aleph	Columna tabla Oracle Aleph	Archivo modificado	Ubicación dentro del archivo modificado	Back end (Encargados de sistemas)	Front end (Usuarios finales)	
Z35 (Events)	Z35_REC_KEY			key_transaccion	ID Transacción	61
Z35 (Events)	Z35_EVENT_TYPE	CatCodigosOracleAleph.docx	APARTADO: TIPO DE TRANSACCIÓN, Desglose de los códigos numéricos para la columna Z35_EVENT_TYPE	tipo_transaccion	Tipo de transacción	Devolución
Z35 (Events)	Z35_EVENT_DATE			dia_transaccion	Día de la transacción	20120221
Z35 (Events)	Z35_EVENT_HOUR			hora_transaccion	Hora de la transacción	1011
		Subdimensión tipo_devolucion		key_tipo_devolucion	Tipo de devolución	Devolución vencida
Cálculo programado en el DW				grupo_transaccion	Agrupación de transacciones	0

Fuente: Elaboración propia.

**Tabla 22. Mapeo de las fuentes de datos de las perspectivas del modelo lógico pertenecientes a la dimensión TIPO\_DEVOLUCION**

Dimensión: Tipo_Devolucion						
Fuentes de los datos				Perspectiva del modelo lógico		Ejemplo de los datos
				Terminología DW		
Tabla Oracle Aleph	Columna tabla Oracle Aleph	Archivo modificado	Ubicación dentro del archivo modificado	Back end (Encargados de sistemas)	Front end (Usuarios finales)	
				key_tipo_devolucion		
Z36H (Loans history)	Z36H_LOAN_DATE			fecha_prestamo_mdoc	Fecha de préstamo	20120215
Z36H (Loans history)	Z36H_DUE_DATE			fecha_vencimiento_mdoc	Fecha de vencimiento	20120220
Z36H (Loans history)	Z36H_RETURNED_DATE			fecha_devolucion_mdoc	Fecha de devolución	20120221
Z36H (Loans history)	Z36H_LOAN_HOUR			hora_prestamo_mdoc	Hora de préstamo	1254
Z36H (Loans history)	Z36H_DUE_HOUR			hora_vencimiento_mdoc	Hora de vencimiento	2100
Z36H (Loans history)	Z36H_RETURNED_HOUR			hora_devolucion_mdoc	Hora de devolución	1011

Fuente: Elaboración propia.

**Tabla 23. Mapeo de las fuentes de datos de las perspectivas del modelo lógico pertenecientes a la dimensión BIBLIOTECARIO**

<b>Dimensión: Bibliotecario</b>						
<b>Fuentes de los datos</b>				<b>Perspectiva del modelo lógico</b>		<b>Ejemplo de los datos</b>
				<b>Terminología DW</b>		
<b>Tabla Oracle Aleph</b>	<b>Columna tabla Oracle Aleph</b>	<b>Archivo modificado</b>	<b>Ubicación dentro del archivo modificado</b>	<b>Back end (Encargados de sistemas)</b>	<b>Front end (Usuarios finales)</b>	
Z35 (Events)	Z35_CATALOGER_NAME			key_bibliotecario	ID Bibliotecario	DSARMANDO
Z35 (Events)	Z35_EVENT_TYPE	CatCodigosOracleAleph.docx	APARTADO: TIPO DE TRANSACCIÓN, Desglose de los códigos numéricos para la columna Z35_EVENT_TYPE: valor 82	error_devolucion	Error de devolución	0

Fuente: Elaboración propia.

**Tabla 24. Mapeo de las fuentes de datos de las perspectivas del modelo lógico pertenecientes a la dimensión *MULTA***

Dimensión: Multa						
Fuentes de los datos				Perspectiva del modelo lógico		Ejemplo de los datos
				Terminología DW		
Tabla Oracle Aleph	Columna tabla Oracle Aleph	Archivo modificado	Ubicación dentro del archivo modificado	Back end (Encargados de sistemas)	Front end (Usuarios finales)	
Z31 (Cash)	Z31_REC_KEY			key_multa	ID Multa	289939 201610111892311
Z31 (Cash)	Z31_STATUS			estado_multa	Estado de la multa	Pagada
Z31 (Cash)	Z31_PAYMENT_CATALOGER			bibliotecario_cobro	Bibliotecario que cobro la multa	DSALBERTO
Z31 (Cash)	Z31-TYPE	CatCodigosOracleAleph.docx	APARTADO: TIPO DE MULTA, Desglose de los códigos numéricos para la columna Z31_TYPE	tipo_multa	Tipo de multa	Late return
Z31 (Cash)	Z31_PAYMENT_DATE_KEY			fecha_hora_pago_multa	Fecha y hora de pago de multa	201610201144

Fuente: Elaboración propia.

**Tabla 25. Mapeo de las fuentes de datos de las perspectivas del modelo lógico pertenecientes a la dimensión TIEMPO**

Dimensión: Tiempo						
Fuentes de los datos				Perspectiva del modelo lógico		Ejemplo de los datos
				Terminología DW		
Tabla(s) Oracle Aleph	Columna tabla Oracle Aleph	Archivo modificado	Ubicación dentro del archivo modificado	Back end (Encargados de sistemas)	Front end (Usuarios finales)	
Z13, Z30H, Z303, Z305, Z35, Z31, Z36H (Fecha y hora de los eventos)	Z13_IMPRINT, SUBSTR(Z13_IMPRINT,-4,4)=AÑO, Z30H_OPEN_DATE, Z30H_DATE_LAST_RETURN, Z30H-H-DATE, Z30H-H-HOUR, Z303_OPEN_DATE, Z303_UPDATE_DATE, Z305_EXPIRY_DATE, Z35_EVENT_DATE, Z35_EVENT_HOUR, Z31_PAYMENT_DATE_KEY, Z36H_LOAN_DATE, Z36H_DUE_DATE, Z36H_RETURNED_DATE, Z36H_LOAN_HOUR, Z36H_DUE_HOUR, Z36H_RETURNED_HOUR			key_tiempo	ID Tiempo	
Z30H, Z303, Z305, Z35, Z31, Z36H (Fecha de las transacciones)	Z30H_OPEN_DATE, Z30H_DATE_LAST_RETURN, Z30H-H-DATE, Z303_OPEN_DATE, Z303_UPDATE_DATE, Z305_EXPIRY_DATE, Z35_EVENT_DATE, Z31_PAYMENT_DATE_KEY, Z36H_LOAN_DATE, Z36H_DUE_DATE, Z36H_RETURNED_DATE			fecha	Fecha	20120221
Z30H, Z35, Z31, Z36H (Hora de las transacciones)	Z30H-H-HOUR, Z35_EVENT_HOUR, Z31_PAYMENT_DATE_KEY, Z36H_LOAN_HOUR, Z36H_DUE_HOUR, Z36H_RETURNED_HOUR			hora	Hora	1011
Cálculo programado en el DW				mes_num	Mes (numérico)	2
Cálculo programado en el DW				mes_text	Mes (textual)	FEBRERO
Cálculo programado en el DW				año	Año	2012
Cálculo programado en el DW				semestre	Semestre	1
Cálculo programado en el DW				semana	Semana	8
Cálculo programado en el DW				dia_habil	Día hábil	MA

Fuente: Elaboración propia.

**Tabla 26. Mapeo de las fuentes de datos de los indicadores del modelo lógico pertenecientes a la tabla de hechos CIRCULACION\_MDOC**

<b>Tabla de hechos: Circulacion_MDoc</b>					
<b>Fuentes de los datos</b>				<b>Indicador del modelo lógico</b>	
				<b>Terminología DW</b>	
<b>Tabla Oracle Aleph</b>	<b>Columna tabla Oracle Aleph</b>	<b>Archivo modificado</b>	<b>Ubicación dentro del archivo modificado</b>	<b>Back end (Encargados de sistemas)</b>	<b>Front end (Usuarios finales)</b>
				key_transaccion	
				key_tiempo	
				key_tipo_transaccion	
				key_mdoc	
				key_usuario	
				key_bibliotecario	
				key_multa	
				key_tipo_devolucion	
Z36H				numero_prestamos	Número de préstamos
Z36H				numero_renovaciones	Número de renovaciones
Z36H				numero_devoluciones	Número de devoluciones
Z35				numero_errores_devolucion_mdoc	Número de errores de devolución de MDoc
Z31				numero_multas	Número de multas
Z31				numero_multas_cobradas	Número de multas cobradas
Z31				numero_multas_canceladas	Número de multas canceladas
Z36H				numero_devoluciones_previas	Número de devoluciones previas
Z36H				numero_devoluciones_en_tiempo	Número de devoluciones en tiempo
Z36H				numero_devoluciones_vencidas	Número de devoluciones vencidas
Z36H				numero_dias_vencidos	

Fuente: Elaboración propia.

### 2.2.2.3 Granularidad

Ya establecido el mapeo de las fuentes de datos se describen a continuación los campos de cada una de las perspectivas a partir de los cuales fue posible analizar los indicadores establecidos en el modelo conceptual. Este apartado hace referencia al nivel de detalle en el que fueron ingresadas (y pueden ser consultadas) las métricas a la tabla de hechos del DW. La tabla de hechos se relaciona al proceso de circulación del material documental, por lo cual las métricas consideradas fueron las siguientes<sup>41</sup>:

TH

#### CIRCULACION\_MDOC

**key transaccion:** Llave primaria de la transacción realizada en el proceso de circulación del MDoc.

**key tiempo:** Llave primaria de la dimensión que define el espacio temporal en el cual ocurrió la transacción.

**key tipo transaccion:** Llave primaria que define el tipo de transacción realizada.

**key mdoc:** Es la llave primaria de la tabla de la dimensión MDOC y representa a un libro específico.

**key usuario:** Llave primaria que identifica al usuario que participa en la transacción.

**key bibliotecario:** Llave primaria que identifica al bibliotecario que participa en la transacción.

**key multa:** Llave primaria de la dimensión de multa por devolución tardía del MDoc.

**key tipo devolucion:** Llave primaria que indica cuál es el tipo de devolución del MDoc prestado: devolución previa, devolución en tiempo o devolución vencida.

**numero prestamos:** El indicador número de préstamos representa el conteo del número de veces que un ítem (MDoc) específico ha sido prestado.

**numero renovaciones:** El indicador número de renovaciones representa el conteo del número de veces que un ítem (MDoc) específico ha sido renovado.

**numero devoluciones:** El indicador número de devoluciones representa el conteo del número de veces que un ítem (MDoc) específico ha sido devuelto.

**numero errores devolucion mdoc:** El indicador número de errores de devolución de MDoc representa el conteo del número de veces que un ítem (MDoc) específico ha sido devuelto sin estar en préstamo.

**numero multas:** El indicador número de multas representa el conteo del número de veces que un ítem (MDoc) específico ha sido prestado y ha sido devuelto después de la fecha especificada para la devolución, por lo cual se genera una transacción de multa.

**numero multas cobradas:** El indicador número de multas cobradas representa el conteo del número de veces que un ítem (MDoc) específico ha generado multas y éstas han sido cobradas.

**numero multas canceladas:** El indicador número de multas canceladas representa el conteo del número de veces que un ítem (MDoc) específico ha generado multas y éstas han sido canceladas.

**numero devoluciones previas:** El indicador número de devoluciones previas representa el conteo del número de veces que un ítem (MDoc) específico ha sido devuelto antes de la fecha de vencimiento establecida por el sistema.

**numero devoluciones en tiempo:** El indicador número de devoluciones en tiempo representa el conteo del número de veces que un ítem (MDoc) específico ha sido devuelto justo en la fecha de vencimiento establecida por el sistema.

---

<sup>41</sup> Los nombres de las métricas y los nombres de los datos que integran cada una de las perspectivas del modelo conceptual se muestran en minúsculas, sin acentos y con sus elementos separados por guiones bajos. Esta es la manera en que estos elementos fueron nombrados en el sistema *MySQL Workbench*, que fue empleado para la estructuración del DW de la BDESSalud. Lo anterior corresponde a un convencionalismo de ingeniería en sistemas que facilita la programación del DW y el manejo de los elementos considerados en éste.

**numero\_devoluciones\_vencidas:** El indicador número de devoluciones vencidas representa el conteo del número de veces que un ítem (MDoc) específico ha sido devuelto luego de la fecha de vencimiento establecida por el sistema.

**numero\_dias\_vencidos:** Es el conteo de los días transcurridos desde la fecha de vencimiento del préstamo del MDoc especificada por el sistema y el día en que el MDoc fue devuelto a la biblioteca.

A continuación, se enlistan cuáles son los datos que integran a cada una de las perspectivas consideradas en el modelo conceptual:

## **MDOC**

**key\_mdoc:** Es la llave primaria de la tabla de la dimensión MDOC y representa a un libro específico.

**key\_tema:** Llave primaria del tema. Hace referencia a la temática del libro o MDoc.

**key\_rec:** Llave que sirve para agrupar los libros que corresponden a un mismo título y autor.

**título:** Título del libro o MDoc.

**autor:** Autor del libro o MDoc.

**pie\_imprensa:** Texto que hace referencia al lugar (país), editorial y año de publicación del MDoc.

**anio\_mdoc:** Año de publicación del MDoc.

**clasificacion:** Código alfanumérico que define el ordenamiento en estantería y temática del MDoc.

**tipo\_coleccion:** Colección a la que fue asignado el MDoc.

**tipo\_material:** Tipo de material a que pertenece el MDoc (libro, disco, DVD, mapa, partitura, etc.)

**tipo\_adquisicion:** Determina la manera en que fue adquirido el MDoc: compra directa, proyecto institucional, donación, canje, etc.

**fecha\_ingreso\_mdoc:** Fecha de ingreso al sistema del MDoc.

**fecha\_ultima\_devolucion\_mdoc:** Fecha de última devolución del MDoc.

**estado\_mdoc:** Estado del MDoc para efectos de préstamo. Se relaciona con el tipo de préstamo realizado, dependiendo del valor señalado: según el número de días en préstamo, el tipo de préstamo, el ítem en préstamo, etc.

**mdoc\_cambios:** Cambios hechos al MDoc.

**mdoc\_ultimo\_cambio:** Último cambio hecho al MDoc.

**fecha\_baja\_mdoc:** Fecha de baja del MDoc del sistema.

**hora\_baja\_mdoc:** Hora de baja del MDoc del sistema.

## **TEMA\_MDOC**

**key\_tema:** Llave primaria del tema. Hace referencia a la temática del libro o MDoc.

**clase\_principal:** Clase principal del Sistema de Clasificación Decimal de Dewey a la cual pertenece el MDoc.

**division\_subclase:** División o subclase del Sistema de Clasificación Decimal de Dewey a la cual pertenece el MDoc.

**seccion\_grupo:** Tema o sección del Sistema de Clasificación Decimal de Dewey a la cual pertenece el MDoc.

## **USUARIO**

**key\_usuario:** Llave primaria que identifica al usuario que participa en la transacción.

**key\_tipo\_usuario:** Llave primaria que indica el tipo de usuario

**key\_carrera:** Llave primaria que identifica a cuál carrera pertenece el usuario.

**key\_unidad\_academica:** Llave primaria que identifica a cuál unidad académica o facultad pertenece el usuario.

**key\_biblioteca\_adscripcion:** Llave primaria que indica a cuál biblioteca está asignado el usuario.

**sexo\_usuario:** Sexo del usuario.



**fecha\_creacion\_reg\_usuario:** Fecha en la cual fue creado el registro del usuario en el sistema.  
**fecha\_actualizacion\_reg\_usuario:** Fecha en la cual fue actualizado el registro del usuario en el sistema.  
**fecha\_vencimiento\_reg\_usuario:** Fecha en la cual vence/venció el registro del usuario en el sistema.

#### **TIPO\_USUARIO**

**key\_tipo\_usuario:** Llave primaria que indica el tipo de usuario.  
**tipo\_usuario:** Tipo de usuario que lleva a cabo la transacción, el cuál puede ser alguna de las variantes de las etiquetas empleadas en Aleph por la BDESSalud: Alumno<sup>42</sup>, Académico, Administrativo, Externo, Préstamo interno, Posgrado o Sistema abierto.

#### **CARRERA**

**key\_carrera:** Llave primaria que identifica cual es la carrera a que pertenece el usuario.  
**carrera:** Carrera a la cual pertenece el usuario.  
**nivel\_academico:** Nivel académico que está siendo cursado por el usuario Alumno.  
**programa\_educativo:** Programa educativo al que pertenece el usuario.

#### **UNIDAD\_ACADEMICA**

**key\_unidad\_academica:** Llave primaria que identifica a cuál unidad académica o facultad pertenece el usuario.  
**unidad\_academica:** Unidad académica a la cual pertenece el usuario.

#### **BIBLIOTECA\_ADSCRIPCION**

**key\_biblioteca\_adscripcion:** Llave primaria que indica a cuál biblioteca está asignado el usuario.  
**biblioteca\_adscripcion:** Biblioteca a la cual pertenece el usuario.

#### **TIPO\_TRANSACCION**

**key\_tipo\_transaccion:** Llave primaria que define el tipo de transacción realizada.  
**d\_tipo\_transaccion:** Transacción realizada.  
**fecha\_transaccion:** Fecha en la cual fue realizada la transacción.  
**hora\_transaccion:** Hora en la cual fue realizada la transacción.  
**grupo\_transaccion:** Agrupación o suma de las transacciones realizadas.

#### **TIPO\_DEVOLUCION**

**key\_tipo\_devolucion:** Llave primaria que indica cuál es el tipo de devolución del MDoc prestado: devolución previa, devolución en tiempo o devolución vencida.  
**fecha\_prestamo\_mdoc:** Fecha en la cual fue prestado el MDoc.  
**fecha\_vencimiento\_mdoc:** Fecha establecida por el sistema en la cual vence el préstamo del MDoc.  
**fecha\_devolucion\_mdoc:** Fecha en la cual fue devuelto el MDoc a la biblioteca.

#### **BIBLIOTECARIO**

**key\_bibliotecario:** Llave primaria que identifica al bibliotecario que participa en la transacción.  
**errordevolucion\_bibliotecario:** Transacción en la cual es devuelto un MDoc que no se encontraba prestado.

---

<sup>42</sup> Se empleó el término Alumno en lugar del término Estudiante para el diseño y estructuración del DW de la BDESSalud y para la realización del ejercicio de bibliominería, debido a que "Alumno" es uno de los tipos de usuario etiquetados en el sistema Aleph por convención en el SUBA.

## **MULTA**

**key\_multa:** Llave primaria de la dimensión de multa por devolución tardía del MDoc.

**estado\_multa:** Estado de la multa o transacción en efectivo: pagada, pendiente y/o cancelada.

**bibliotecario\_cobro:** Bibliotecario que llevó a cabo el cobro de la multa.

**tipo\_multa:** Código de la multa o transacción en efectivo: devolución tardía, reclamo de devolución, solicitud de reserva, etc.

**fecha\_hora\_pago\_multa:** Fecha y hora en las cuales fue pagada la multa.

## **TIEMPO**

**key\_tiempo:** Llave primaria de la dimensión que define el espacio temporal en el cual ocurrió la transacción.

**fecha:** Fecha en la cual se llevó a cabo la transacción.

**hora:** Hora en la cual se llevó a cabo la transacción.

**mes\_num:** Mes en el cual fue llevada a cabo la transacción (expresado en número).

**mes\_text:** Mes en el cual fue llevada a cabo la transacción (expresado en texto).

**año:** Año en el cual fue llevada a cabo la transacción.

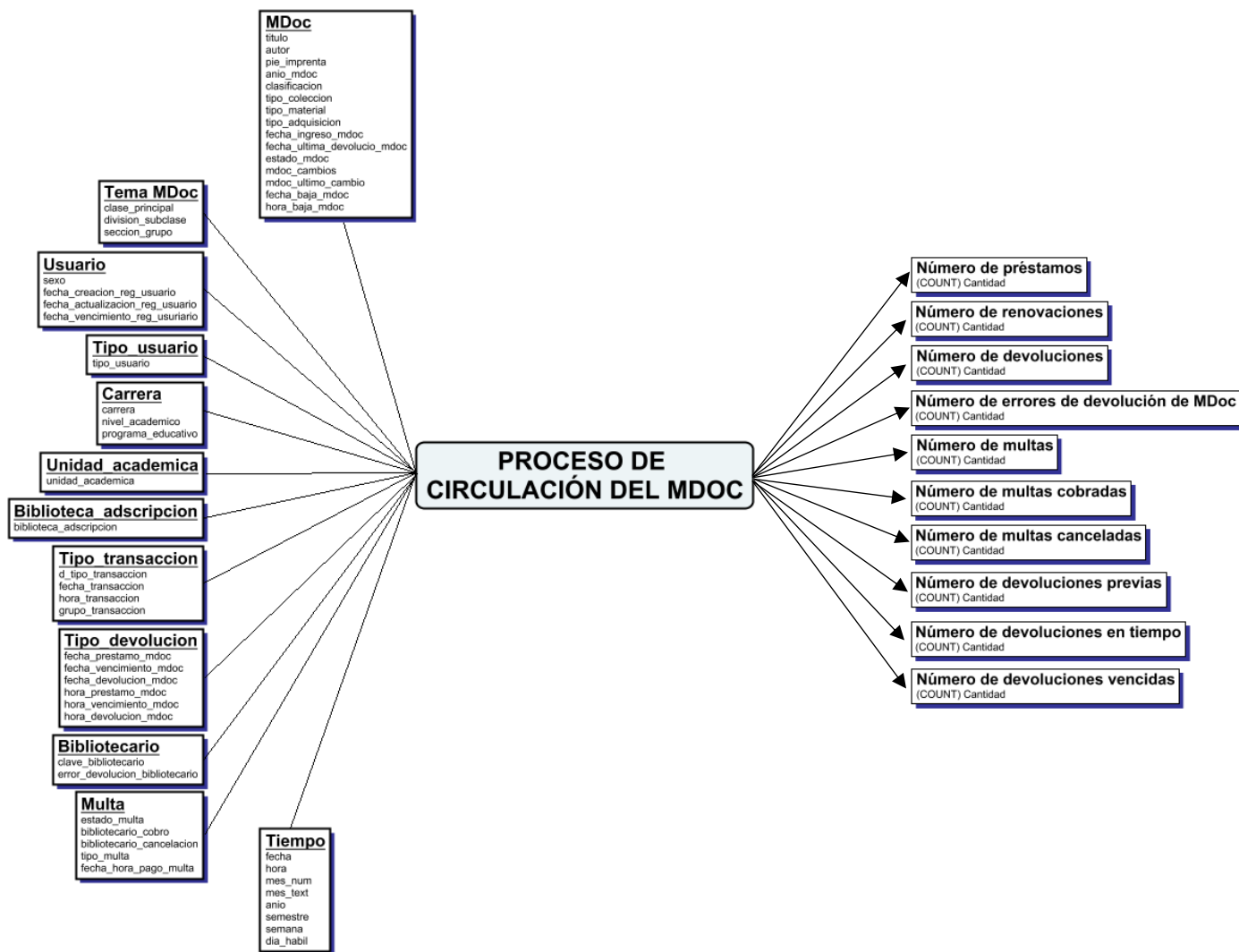
**semestre:** Semestre en el cual fue llevada a cabo la transacción.

**semana:** Semana en la cual fue llevada a cabo la transacción.

**dia\_habil:** Día hábil en el cual fue llevada a cabo la transacción.

### ***2.2.2.4 Modelo conceptual ampliado del Data warehouse***

Esta etapa implica la descripción gráfica de los resultados obtenidos en las etapas previas: el modelo conceptual se amplía al colocar debajo de cada perspectiva los datos o atributos que la conformarán y debajo de cada indicador su respectiva fórmula de cálculo, tal como se muestra en la Figura 9.



**Figura 9. Modelo conceptual ampliado del DW de la BDESSalud**

Fuente: Elaboración propia.

### 2.2.3 Modelo lógico del data warehouse

De acuerdo con la metodología Hefesto, la siguiente etapa es la confección del modelo lógico de la estructura del DW, considerando el modelo conceptual creado con anterioridad. Un modelo lógico es la representación de una estructura de datos que puede procesarse y almacenarse en algún sistema de gestión de bases de datos (SGBD) (Bernabeu y García, 2017, p. 144).

### 2.2.3.1 Tipología

El tipo de esquema adoptado para la estructuración del DW fue el de copo de nieve, pues la dimensión Usuario fue implementada con más de una tabla de datos, mismas que fueron designadas como subdimensiones<sup>43</sup>:

**Dimensión:** Usuario

**Subdimensión:** Tipo de usuario

**Subdimensión:** Carrera

**Subdimensión:** Unidad académica

**Subdimensión:** Biblioteca de adscripción

La finalidad de implementar tales subdimensiones fue la de normalizar la base de datos para eliminar redundancias y así optimizar el espacio de almacenamiento de los datos. Se consideró asimismo el inconveniente de que este tipo de esquema puede requerir más tiempo para realizar las consultas al DW por el incremento en el número de uniones (*joins*), sin embargo, el tiempo de respuesta del DW no es un factor crítico para los usuarios de éste, por lo cual la tipología puede considerarse como apropiada para el DW de la BDESSalud.

### 2.2.3.2 Tablas de dimensiones

El siguiente paso de la metodología Hefesto es el diseño de las tablas de dimensiones<sup>44</sup> que conformarán el DW, “para lo cual se debe de tomar en cuenta cual será el nombre de la tabla, añadir el campo que será la clave primaria y los respectivos nombres a los campos” (Haro y Pérez, 2014, p. 112). En este proyecto se emplearon las siguientes dimensiones y subdimensiones:

**Dimensión:** MDoc

---

<sup>43</sup> Esto ocurrió también con la dimensión MDoc, pero en este caso únicamente necesitó la implementación de una subdimensión: Tema MDoc.

<sup>44</sup> También conocidas como catálogos (de perspectivas o variables).

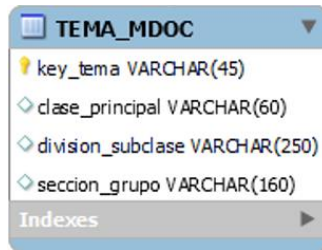
La tabla de dimensión MDoc se refiere a los ítems, libros o materiales que conforman las colecciones en formato físico de la BDESSalud. Tiene la llave primaria que funge como identificador único de cada libro, así como los campos señalados en la Figura 10.



**Figura 10. Dimensión MDoc**

Fuente: Elaboración propia.

**Subdimensión:** Tema MDoc. Esta subdimensión hace referencia a la temática del libro o MDoc. Tiene la llave primaria del tema, a partir de la cual es posible determinar la clase principal, la subdivisión o subclase, así como la sección o grupo temático a que corresponde el MDoc, lo cual se muestra en la Figura 11.



**Figura 11. Subdimensión tema MDoc**

Fuente: Elaboración propia.

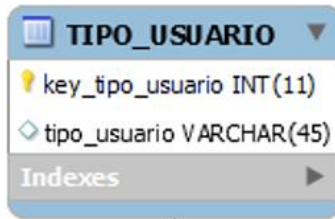
**Dimensión:** Usuario. Esta dimensión abarca la información relativa al usuario que participó en las transacciones del proceso de circulación. Maneja la llave primaria que identifica al usuario, así como los elementos señalados en la Figura 12.



**Figura 12. Dimensión usuario**

Fuente: Elaboración propia.

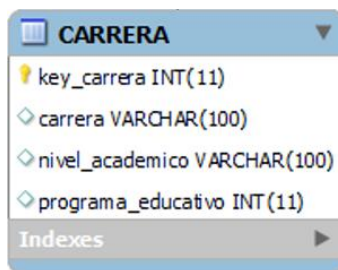
**Subdimensión:** Tipo de usuario. Indica qué tipo de usuario realizó la transacción: Alumno, Académico, Administrativo, Externo, etc. Sus elementos son visibles en la Figura 13.



**Figura 13. Subdimensión tipo usuario**

Fuente: Elaboración propia.

**Subdimensión:** Carrera. Esta subdimensión identifica la carrera a la cual pertenece el usuario que llevó a cabo la transacción. Los elementos que la componen se listan en la Figura 14.



**Figura 14. Subdimensión carrera**

Fuente: Elaboración propia.

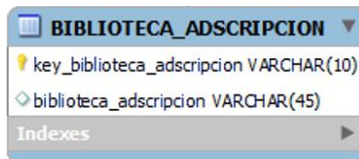
**Subdimensión:** Unidad académica. Esta subdimensión señala a cuál unidad académica o facultad pertenece el usuario que llevó a cabo la transacción. Los elementos que la componen se listan en la Figura 15.



**Figura 15. Subdimensión unidad académica**

Fuente: Elaboración propia.

**Subdimensión:** Biblioteca de adscripción. Señala a qué biblioteca está asignado el usuario que llevó a cabo la transacción. La Figura 16 indica sus componentes.



**Figura 16. Subdimensión biblioteca de adscripción**

Fuente: Elaboración propia.

**Dimensión:** Tipo de transacción. Establece cuál fue la transacción efectuada. Los elementos de esta dimensión se muestran en la Figura 17.

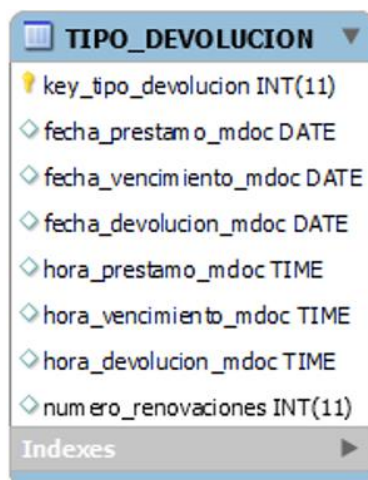


**Figura 17. Dimensión tipo de transacción**

Fuente: Elaboración propia.



**Dimensión:** Tipo de devolución. Sirve para determinar cuál fue el tipo de devolución del MDoc prestado: devolución previa, devolución en tiempo o devolución vencida. Para ello cuenta con los elementos señalados en la Figura 18.



**Figura 18. Dimensión tipo de devolución**

Fuente: Elaboración propia.

**Dimensión:** Bibliotecario. Esta dimensión sirve para identificar al bibliotecario que participó en la transacción. Sus elementos se listan en la Figura 19.



**Figura 19. Dimensión bibliotecario**

Fuente: Elaboración propia.

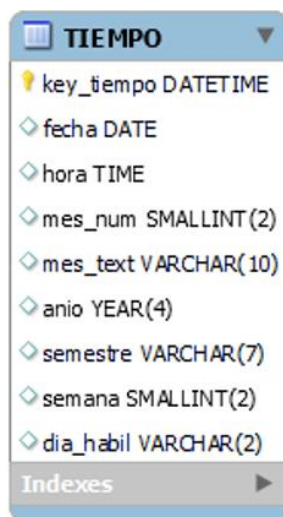
**Dimensión:** Multa. En esta dimensión está el registro de las devoluciones tardías del MDoc que generaron multa. Sus componentes se muestran en la Figura 20.



**Figura 20. Dimensión multa**

Fuente: Elaboración propia.

**Dimensión:** Tiempo. Esta dimensión define el espacio temporal en el cual ocurrió la transacción. Los elementos considerados en esta dimensión se muestran en la Figura 21.




**Figura 21. Dimensión tiempo**

Fuente: Elaboración propia.

### 2.2.3.3 Tabla de hechos

Las dimensiones y subdimensiones anteriormente referidas alimentaron la tabla de hechos del DW de la BDESSalud, la cual está relacionada al proceso de circulación del MDoc en formato físico. La tabla de hechos de este proyecto se muestra en la Figura 22.



Field Name	Data Type
key_tiempo	DATETIME
key_tipo_transaccion	INT(11)
key_mdoc	VARCHAR(45)
key_usuario	VARCHAR(20)
key_bibliotecario	VARCHAR(45)
key_multa	VARCHAR(45)
key_tipo_devolucion	INT(11)

Indexes

**Figura 22. Tabla de hechos del proceso de circulación del MDoc**

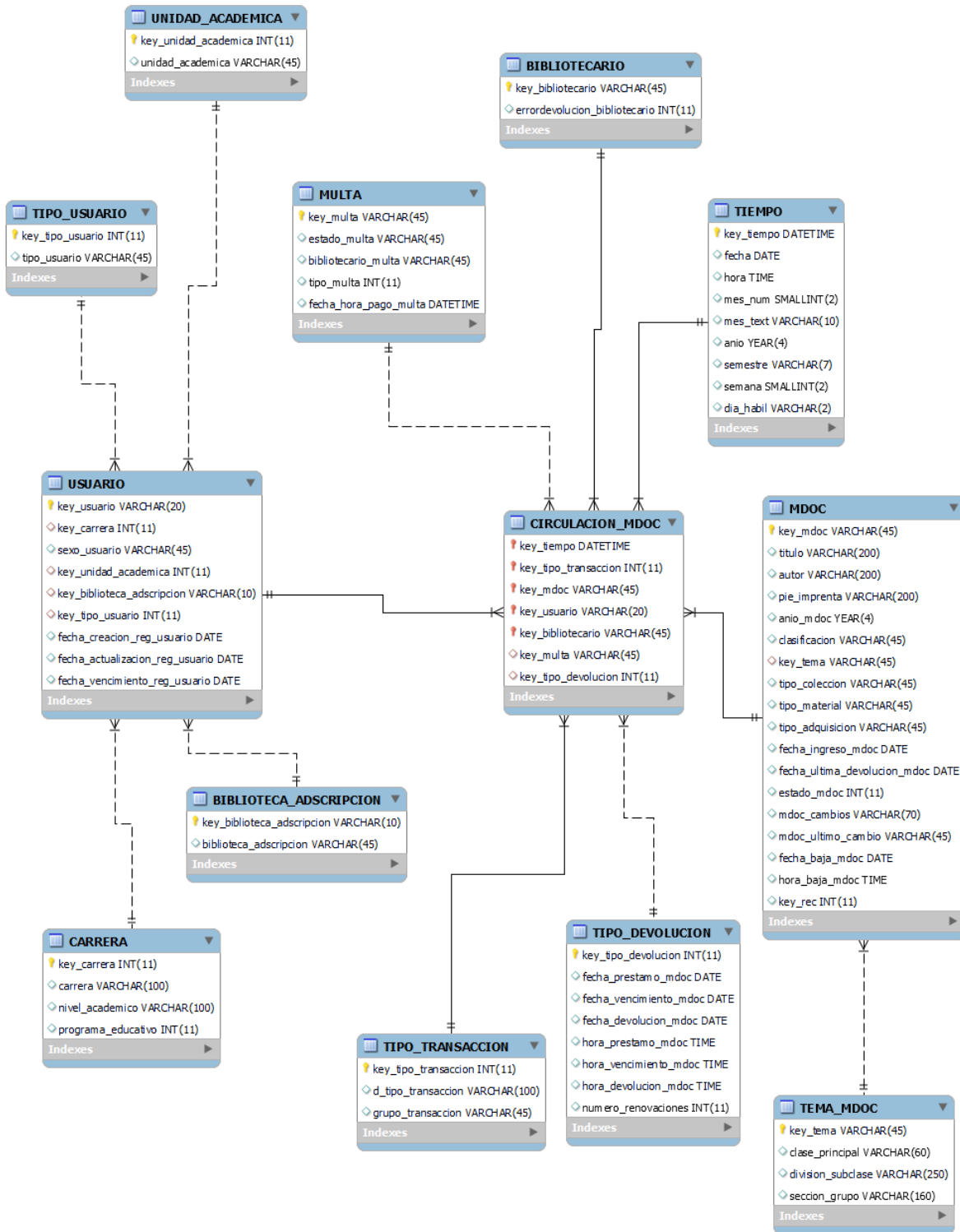
Fuente: Elaboración propia.

Las métricas o indicadores correspondientes a la tabla de hechos de Circulacion\_MDoc son:

- Número de préstamos
- Número de renovaciones
- Número de devoluciones
- Número de errores de devolución del MDoc
- Número de multas
- Número de multas cobradas
- Número de multas canceladas
- Número de devoluciones previas
- Número de devoluciones en tiempo
- Número de devoluciones vencidas

#### **2.2.3.4 Uniones**

Luego de la definición de las dimensiones y de la tabla de hechos del DW de la BDESSalud se procedió, tal como se recomienda en la metodología Hefesto, a determinar las uniones o relaciones correspondientes entre las tablas del modelo, lo cual dio como resultado el diagrama presentado en la Figura 23.



**Figura 23. Uniones de las tablas dimensionales y subdimensionales con la tabla de hechos del proceso de circulación del MDoc de la BDESSalud**

Fuente: Elaboración propia.

Es importante mencionar que para la estructuración del DW de la BDESSalud se contó con el apoyo de un Ingeniero en Sistemas que se encargó de los aspectos de programación para crear la estructura del DW, así como la etapa de ETL e implementación de la interfaz de consulta y visualización del DW, así como de la asesoría global de un maestro en Ciencias de la Computación, quien supervisó todo el proyecto y apoyó en el ejercicio de bibliominería del DW de la BDESSalud.

## **2.2.4 Integración de los datos**

“Una vez construido el modelo lógico, se deberá [poblar] con datos” (Bernabeu y García, 2017, p. 152). Como veremos en los siguientes apartados, en esta etapa de la construcción del DW se llevaron a cabo labores como la obtención de los datos necesarios, la aplicación de técnicas de limpieza de datos, control de calidad de datos y procesos de extracción, transformación y carga (*ETL*, por sus siglas en inglés). Para ello fue necesario definir las políticas y estrategias para la carga inicial del DW, así como de su proceso de actualización a futuro<sup>45</sup>. A continuación, se presenta una breve descripción del software empleado para la confección del DW de la BDESSalud.

### **2.2.4.1 Programa informático para la estructuración del DW: MySQL Workbench**

El programa seleccionado para la estructuración del DW de esta tesis fue *MySQL Workbench*, que se describe como una “herramienta oficial de interfaz gráfica de usuario (*GUI*, por sus siglas en inglés) para *MySQL*. [El programa] permite diseñar, crear y explorar [los] esquemas de bases de datos, trabajar con objetos de base de datos e insertar datos, así como diseñar y ejecutar consultas SQL para trabajar con los datos almacenados. También puede migrar esquemas y datos de otros proveedores de bases de datos a [la base de datos creada en] *MySQL*”. *MySQL* es un sistema de administración

---

<sup>45</sup> Véanse: Anexo 7. Políticas de carga para el llenado del DW BDESSalud y Anexo 8. Políticas y estrategias de actualización del DW BDESSalud.

o gestión de bases de datos relacionales de código abierto respaldado por Oracle y fundamentado en *SQL*. Funciona bajo un modelo cliente-servidor. En el caso de esta tesis, la carga de las tablas extraídas del SGIB se realizó en primera instancia en una base de datos transicional diseñada en *MySQL Workbench* que permitió llevar a cabo las labores de exploración inicial de los datos, así como la transformación de estos para una mejor integración en el DW. Todo ello fue a partir de una infraestructura creada en la nube y no en un servidor físico. La decisión de emplear este programa estribó sobre todo en que es un programa de acceso abierto y su uso está relativamente generalizado entre los administradores de sistemas, incluido el actual jefe de la Unidad de Sistemas del SUBA, lo cual puede favorecer en el seguimiento, uso y actualización del DW de la BDESSalud. En el siguiente apartado se indica cómo fue llevada a cabo la carga inicial de los datos, así como cuáles fueron los procedimientos empleados para su limpieza, transformación y carga.

#### **2.2.4.2 Proceso ETL**

##### Extracción u obtención del conjunto inicial de datos

El análisis meta de la tesis se fundamentó principalmente en la información resguardada en una base de datos transaccional: el sistema integrado de gestión bibliotecaria Aleph, el cual alberga los datos referentes a las operaciones de la BDESSalud en un sistema relacional configurado en tablas de Oracle. En principio se había considerado emplear los datos derivados de los reportes de LogAleph para la construcción del DW. Sin embargo, hubo variables de interés que no pudieron ser obtenidas por esta vía debido a que los reportes dónde estaban contenidas presentaron fallos para su obtención. Además, la posibilidad de enlazar las distintas variables entre sí era muy limitada por la carencia de llaves primarias y/o secundarias válidas. Por lo anterior, se solicitó al jefe de la Unidad de Sistemas del SUBA apoyo para sustraer la información necesaria desde las tablas Oracle del Aleph en las cuales se contenían los atributos necesarios para el diseño del DW. Así, los datos recabados corresponden a los generados en el periodo comprendido entre el 01 de enero de 2010 al 31 de diciembre de 2019. Para obtenerlos

se solicitó al jefe de la Unidad de Sistemas la **extracción** (a partir de consultas SQL) de las tablas de interés y las columnas que contenían las variables necesarias para la estructuración inicial del DW de la BDESSalud. De acuerdo con los requerimientos de datos planteados por el diseño previo del DW, las tablas empleadas fueron las siguientes<sup>46</sup>:

- Z13 (Registro bibliográfico corto)
- Z30 (Ítem)
- Z30H (Historial del ítem)
- Z31 (Multas)
- Z35 (Eventos)
- Z36H (Historial de préstamos)
- Z303 (Información global del usuario)
- Z305 (Información local del usuario)
- Z308 (ID de usuario)

Un ejemplo de las consultas en SQL realizadas es la referente a la tabla Z36H, la cual contiene información del historial de préstamos de la BDESSalud:

```
select Z36H_TIME, Z36H_REC_KEY, Z36H_ID, Z36H_NUMBER, Z36H_MATERIAL,  
Z36H_SUB_LIBRARY, Z36H_STATUS,  
  
Z36H_LOAN_DATE, Z36H_LOAN_HOUR, Z36H_DUE_DATE, Z36H_DUE_HOUR,  
Z36H_RETURNED_DATE,  
  
Z36H_RETURNED_HOUR, Z36H_ITEM_STATUS, Z36H_BOR_STATUS,  
Z36H_NO_RENEWAL, Z36H_LOAN_CATALOGER_NAME,  
  
Z36H_RETURN_CATALOGER_NAME, Z36H_RENEW_CATALOGER_NAME,  
Z36H_RENEW_MODE, Z36H_BOR_TYPE
```

---

<sup>46</sup> Para mayor especificidad de las tablas y columnas Aleph Oracle empleadas, véase el apartado 2.3.2.2 Mapeo (de las fuentes de datos).



De manera similar fueron delimitadas las consultas para la extracción del resto de las tablas y columnas de interés. Así, en esta primera aproximación al uso del DW en la BDESSalud se realizó una **extracción inicial**<sup>47</sup> que implicó el uso de los datos contenidos en las tablas proporcionadas por la Unidad de Sistemas, las cuales fueron cargadas en una base de datos provisional en *MySQL Workbench* que fue de suma utilidad para los procesos de exploración, limpieza y transformación de los datos, previos a la subida de los datos en el DW.

#### Transformación: procedimiento de trabajo con los datos

Ya contenida la información de las tablas y columnas de Oracle del Aleph en la tabla provisional, se procedió a explorar la configuración de los datos obtenidos para conocer sus características y detectar valores anómalos o datos faltantes. De acuerdo con Bernabeau y García (2017, p. 153) “[s]e debe evitar que el DW sea cargado con [...] valores faltantes, [...] datos anómalos o faltos de integridad; se deben establecer condiciones y restricciones para asegurar que sólo se utilicen los datos de interés”. La tabla provisional fue útil para la aplicación de *scripts SQL*<sup>48</sup> que sirvieron para la limpieza de los datos cargados. La intención de la implementación de esta fase fue acondicionar los datos para que todos tuviesen el formato adecuado y la coherencia necesaria antes de ser ingresados al DW, además de facilitar la carga al estar al mismo nivel, pero en un diferente esquema no relacional. Para la integración de los datos de las tablas de Excel obtenidas con las consultas hechas al Aleph se procedió a seguir una serie de pasos:

- Se verificó que la codificación de los datos tuviese el formato UTF-8, el cual hace referencia a la integración de acentos y caracteres especiales como lo es la letra “ñ”. En los casos en los cuales no fue aplicada esta codificación se procedió a su búsqueda e identificación para reemplazar los caracteres proporcionados por los correctos.

---

<sup>47</sup> Manual, no automatizada, debido a las limitaciones de acceso directo a la base de datos transaccional.

<sup>48</sup> En su versión 5.7.36. La elección de este lenguaje estribó en que permite manipular de manera sencilla los datos estructurados en una base de datos relacional, lo cual asegura una gestión eficiente y facilita el proceso de *ETL* entre la base transaccional y el DW.

- Además, a partir del uso de funciones de Excel, se procedió a agregar a los identificadores de caracteres de SQL (“, ’, `) la señalización de la integración de éstos como parte del contenido, usando la diagonal invertida antes de ellos.
- Fue necesario ingresar la palabra reservada *NULL*<sup>49</sup> en el caso de las celdas que se encontraron sin información, de manera que todas las filas tuviesen el mismo número de valores y así se agilizara la carga de las tablas transicionales.
- En la Tabla 27 se muestra el análisis realizado de los tipos de datos esperados de acuerdo con la funcionalidad que tendrían dentro del DW, así como con el tamaño que les sería asignado.

**Tabla 27. Tipos de datos, características y funcionalidades**

<b>Tipos de datos</b>	<b>Características restrictivas</b>	<b>Funcionalidades</b>
<b>VARCHAR</b>	Caracteres alfanuméricos.	Permite únicamente búsquedas de cadenas o subcadenas.
<b>INT</b>	Caracteres numéricos enteros, positivos y negativos, con longitud menor a 10 caracteres.	Permite búsquedas de cadenas o subcadenas, además de rangos.
<b>DOUBLE</b>	Caracteres numéricos enteros, positivos y negativos, con longitud menor a 10 caracteres.	Permite las búsquedas de cadenas o subcadenas, así como de rangos numéricos.
<b>DATE</b>	Tipo de dato que representa valores temporales. Formato YYYY-MM-DD.	Permite búsquedas de cadenas o subcadenas (solo fecha, año, mes, día) además de búsquedas en rangos de tiempo.
<b>DATETIME</b>	Tipo de dato que representa valores que contienen partes de fecha y hora. Formato YYYY-MM-DD hh:mm:ss.	Permite búsquedas de cadenas o subcadenas (solo fecha, año, mes, día u hora), además de búsquedas en rangos de tiempo.
<b>SMALLINT</b>	Caracteres numéricos enteros, positivos y negativos con longitud de cinco caracteres.	Permite búsquedas de cadenas o subcadenas, además de rangos numéricos.
<b>TIME</b>	Tipo de dato que representa valores de hora. Formato hh:mm:ss.	Permite búsquedas de cadenas o subcadenas (solo hora, minutos), además de búsquedas en rangos de tiempo.
<b>YEAR</b>	Tipo de dato que representa valores de año. Formato YYYY. Longitud de cuatro caracteres.	Permite las búsquedas de cadenas o subcadenas, así como de rangos numéricos.

Fuente: Elaboración propia.

<sup>49</sup> Campos nulos o inválidos debido a la falta de datos.

- Cuando se ubicaron valores específicos que no cumplían con las restricciones del tipo de dato se procedió a realizar la limpieza y/o transformación de estos datos de manera individual<sup>50</sup>. Asimismo, se procedió a descartar aquellos valores en los cuales no se contó con la información completa para la deducción del dato correcto.
- Posteriormente se generó una tabla nueva en el diagrama transaccional cuyas columnas y valores cumplieron el análisis de datos descrito anteriormente.
- En seguida se adecuaron los datos de entrada en formato de valores separados por comas (CSV, por sus siglas en inglés) y se agregaron comillas a los valores alfanuméricos, completando con estos datos el *script* que se utilizó para ingresarlos al modelo transicional previo a su carga al DW.

Aparte de las tablas obtenidas mediante consultas *SQL* al Aleph, el DW también fue configurado empleando archivos modificados en Excel que sirvieron para delimitar los temas a que pertenece el MDoc, así como para designar la Unidad Académica de los usuarios, y al menos un documento en Word en el cual se muestra un catálogo de códigos para los valores que aparecen en las tablas Oracle Aleph del SUBA<sup>51</sup>.

### DETERMINACIÓN DE LOS TEMAS DEL MDOC

En lo referente a la determinación de los temas del MDoc, en principio se consideró obtener los datos del Aleph referentes a la etiqueta MARC 650 (que indica el encabezamiento temático de materia) que se almacenan en la tabla Z01 (encabezados de acceso). Sin embargo, luego de una revisión realizada por el jefe de la Unidad de Sistemas del SUBA se determinó que sería complejo obtenerlos, además de que no todos los registros bibliográficos contaban con los campos y los datos requeridos. Por lo anterior, para contar con los temas del MDoc se procedió a tomar los datos de

---

<sup>50</sup> A este proceso de transformación de los datos se le conoce también como conversión, casteo (del inglés *cast*, moldear) o modelado de datos específicos para aplicarles el formato requerido por el DW. Se realiza a partir de *scripts SQL*, que suelen ser implementados al mismo tiempo que la extracción por su correlación con el dato original.

<sup>51</sup> Véase el Anexo 9. Catálogo de códigos para valores de las tablas Oracle Aleph del SUBA.

clasificación del último reporte de Inventario de MDoc de la BDESSalud<sup>52</sup>, el cual fue obtenido desde el LogAleph en formato Excel. A partir de aplicación de funciones como ORDENAR y ENCONTRAR de Excel se limpiaron los duplicados y se extrajeron índices de clasificación del MDoc eliminando las notaciones internas<sup>53</sup>. Posteriormente se delimitaron las clases, divisiones y secciones a que pertenecen las clasificaciones del MDoc, para lo cual se empleó la notación del Sistema de Clasificación Decimal Dewey. Esto permitió crear una tabla en formato xlsx que fue incluida en el diseño del DW para permitir asignar temas a los ítems de la colección de la BDESSalud, de la cual se muestra un extracto en la Tabla 28.

---

<sup>52</sup> En específico se descargó el *Reporte 15. Inventario por tipo de material documental*, de la sección de Uso del Acervo del LogAleph.

<sup>53</sup> En Excel, se tomó la columna de clasificación y se pegó en una hoja en blanco. Luego en la opción de *Datos*, en el submenú de *Herramientas de datos* se eligió la opción de *Quitar duplicados* (seleccionando toda la columna). Se encontraron 12,048 valores duplicados, quedaron 7258 valores únicos permanentes. Posteriormente se empleó una serie de fórmulas en Excel para depurar la columna y dejar únicamente los códigos numéricos de la clasificación Dewey. A partir de esto, se encontraron 5,817 valores duplicados y quedaron 1,542 valores únicos.

**Tabla 28. Clasificación – Desagregado de temas para el MDoc de la BDESSalud**

	Clase (SCDD)	División (SCDD)	Sección (SCDD)
Extracto de la clasificación	Clase principal	División o subclase	Sección o grupo
610	Technology	Medicine & health	Medicine & health
610	Technology	Medicine & health	Medicine & health
610.972	Technology	Medicine & health	Medicine & health
611	Technology	Medicine & health	Human anatomy, cytology, histology
611.002	Technology	Medicine & health	Human anatomy, cytology, histology
611.003	Technology	Medicine & health	Human anatomy, cytology, histology
612	Technology	Medicine & health	Human physiology
612.002	Technology	Medicine & health	Human physiology
612.014	Technology	Medicine & health	Human physiology
613	Technology	Medicine & health	Personal health and safety
613.001	Technology	Medicine & health	Personal health and safety
613.02	Technology	Medicine & health	Personal health and safety
614	Technology	Medicine & health	Forensic medicine; incidence of injuries, wounds, disease; public preventive medicine
614.012	Technology	Medicine & health	Forensic medicine; incidence of injuries, wounds, disease; public preventive medicine
614.018	Technology	Medicine & health	Forensic medicine; incidence of injuries, wounds, disease; public preventive medicine
615	Technology	Medicine & health	Pharmacology and therapeutics
615.007	Technology	Medicine & health	Pharmacology and therapeutics
615.1	Technology	Medicine & health	Pharmacology and therapeutics
616	Technology	Medicine & health	Diseases
616.007	Technology	Medicine & health	Diseases
616.008	Technology	Medicine & health	Diseases
617	Technology	Medicine & health	Surgery, regional medicine, dentistry, ophthalmology, otology, audiology
617.001	Technology	Medicine & health	Surgery, regional medicine, dentistry, ophthalmology, otology, audiology
617.002	Technology	Medicine & health	Surgery, regional medicine, dentistry, ophthalmology, otology, audiology
618	Technology	Medicine & health	Gynecology, obstetrics, pediatrics, geriatrics
618.01	Technology	Medicine & health	Gynecology, obstetrics, pediatrics, geriatrics
618.02	Technology	Medicine & health	Gynecology, obstetrics, pediatrics, geriatrics
619	Technology	Medicine & health	Medicine & health [Unassigned]
619.92	Technology	Medicine & health	Medicine & health [Unassigned]
619.92	Technology	Medicine & health	Medicine & health [Unassigned]
641	Technology	Home & family management	Food and drink
641.013	Technology	Home & family management	Food and drink
641.03	Technology	Home & family management	Food and drink
642	Technology	Home & family management	Meals and table service
642.5	Technology	Home & family management	Meals and table service
649	Technology	Home & family management	Child rearing; home care of people with disabilities and illnesses
649	Technology	Home & family management	Child rearing; home care of people with disabilities and illnesses
649.1	Technology	Home & family management	Child rearing; home care of people with disabilities and illnesses
1000	Thesis	Thesis	Thesis
2000	Cubicle	Cubicle	Cubicle

Fuente: Elaboración propia.

Cabe mencionar que debido a las limitaciones temporales no fue posible llevar a cabo la traducción de las asignaciones temáticas, por lo cual los términos aparecen en idioma inglés en la interfaz de consulta y visualización del DW de la BDESSalud. Asimismo, se incluyeron las asignaciones 1000 y 2000 (no previstas en el SCDD) para distinguir a las transacciones referentes a tesis y cubículos en el modelo de circulación de la BDESSalud e incluir éstas en el proceso de circulación.

## ASIGNACIÓN DE UNIDAD ACADÉMICA (FACULTAD)

En la exploración de los datos contenidos en las tablas Oracle del Aleph fue posible verificar que el sistema no cuenta con un apartado en el cual se especifique la unidad académica a que pertenece el usuario. Por ello se consideró utilizar la columna Z303\_FIELD\_3, la cual es empleada por el SUBA para indicar la clave de la carrera a que pertenece el usuario. Los datos de dicha columna se cruzaron con los de una tabla en Excel proporcionada por la jefa de servicios bibliotecarios de la BDESSalud, en la cual se encuentra el desglose de las claves de carrera y se brinda información sobre la ID de unidad, nombre completo de la unidad académica o facultad, la descripción del nivel académico y la descripción del programa educativo. A continuación, se muestra un extracto de la tabulación que sirvió para indicar en el DW cuál es la unidad académica, así como el nivel académico y el programa educativo a que pertenecen los usuarios de la BDESSalud en la Tabla 29.

**Tabla 29. Extracto de la tabla para la asignación de unidad académica (facultad), nivel académico y programa educativo del usuario**

ID de unidad	Unidad académica	Clave de carrera por usuario	Nivel académico	Programa educativo
5200	FACULTAD DE MEDICINA Y CIENCIAS BIOMEDICAS	520004	ESPECIALIDAD	ESPECIALIDAD EN CIRUGIA GENERAL
5200	FACULTAD DE MEDICINA Y CIENCIAS BIOMEDICAS	520020	LICENCIATURA	INGENIERO BIOMEDICO
5200	FACULTAD DE MEDICINA Y CIENCIAS BIOMEDICAS	520021	LICENCIATURA	LICENCIADO EN SALUD PUBLICA
5200	FACULTAD DE MEDICINA Y CIENCIAS BIOMEDICAS	520019	LICENCIATURA	LICENCIADO EN TERAPIA FISICA Y REHABILITACION
5200	FACULTAD DE MEDICINA Y CIENCIAS BIOMEDICAS	520001	LICENCIATURA	MEDICO CIRUJANO Y PARTERO
5300	FACULTAD DE ENFERMERIA Y NUTRIOLOGIA	530001	LICENCIATURA	LICENCIADO EN ENFERMERIA
5300	FACULTAD DE ENFERMERIA Y NUTRIOLOGIA	530002	LICENCIATURA	LICENCIADO EN NUTRICION
5300	FACULTAD DE ENFERMERIA Y NUTRIOLOGIA	530009	MAESTRIA	MAESTRIA EN ENFERMERIA
5300	FACULTAD DE ENFERMERIA Y NUTRIOLOGIA	530011	MAESTRIA	MAESTRIA EN NUTRICION HUMANA
5300	FACULTAD DE ENFERMERIA Y NUTRIOLOGIA	530005	MAESTRIA	MAESTRIA EN SALUD EN EL TRABAJO
5300	FACULTAD DE ENFERMERIA Y NUTRIOLOGIA	5300	ADMINISTRATIVO	PERSONAL ADMINISTRATIVO
5200	FACULTAD DE MEDICINA Y CIENCIAS BIOMEDICAS	5200	ADMINISTRATIVO	PERSONAL ADMINISTRATIVO
5300	FACULTAD DE ENFERMERIA Y NUTRIOLOGIA	530000	ACADEMICO	PERSONAL ACADEMICO
5200	FACULTAD DE MEDICINA Y CIENCIAS BIOMEDICAS	520000	ACADEMICO	PERSONAL ACADEMICO
5300	FACULTAD DE ENFERMERIA Y NUTRIOLOGIA	5300	ACADEMICO	PERSONAL ACADEMICO
5200	FACULTAD DE MEDICINA Y CIENCIAS BIOMEDICAS	5200	ACADEMICO	PERSONAL ACADEMICO

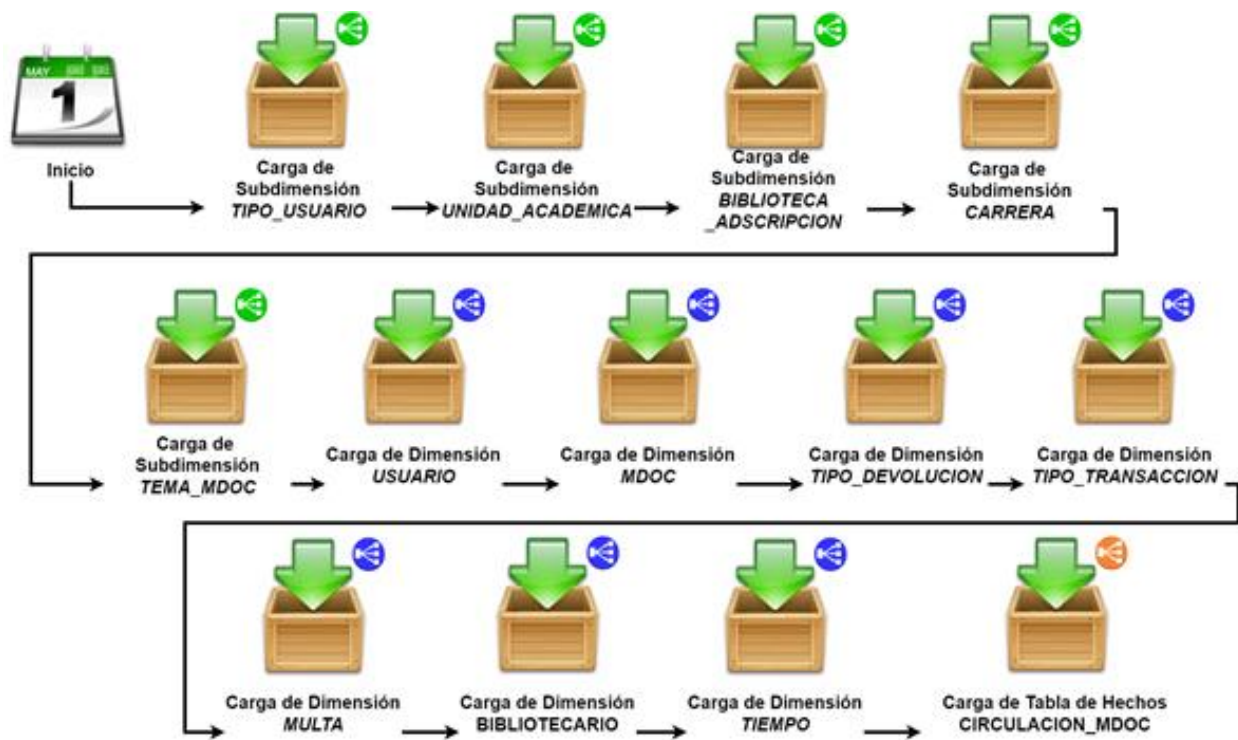
Fuente: Elaboración propia.

## Carga de los datos

Una vez configuradas las tablas transicionales, en las cuales ocurrió la mayoría de las operaciones de limpieza y transformación de los datos, se inicia el proceso de extracción de los datos que no necesitan ser calculados y que son requeridos por el DW. Para ello, se ubicaron los espacios específicos donde se encuentran alojados los datos de interés en las transicionales. Dentro del proceso de extracción se aplicaron las condiciones en donde se eligieron qué datos eran de interés, para evitar la recolección de datos duplicados o de información que ya se encontraba cargada.

En los casos en que no todos los datos fueron obtenidos de la misma tabla se usaron dos opciones para hacer la carga completa: la primera opción (que suele ser la más utilizada en este tipo de procesos) consistió en unir todas las fuentes de datos a partir de llaves únicas que las relacionan, con lo cual se crearon pseudo tablas temporales de donde se extrajeron los datos para ser ingresados. Esta aplicación es a partir de un proceso lineal que aumenta su complejidad dependiendo del tamaño de las tablas, por lo cual su uso no es recomendable cuando se utilizan tablas de tamaño excesivo. La segunda opción fue la carga por tablas, en la cual cada una de ellas es ingresada de manera individual. La primera tabla es ingresada sin comparaciones, mientras que las siguientes serán actualizaciones de la ya creada, realizando la agregación de datos en donde se encuentren las llaves únicas de relación. La utilización de este proceso debe ser tomado en cuenta al momento de realizar las extracciones para evitar variabilidad de los datos.

En la Figura 24 se muestra el proceso a partir del cual se realizó el proceso de carga de las tablas consideradas en el modelo del DW a la estructura creada del DW en *MySQL Workbench*.



**Figura 24. Proceso ETL principal planteado para la carga inicial del DW BDESSalud**

Fuente: Elaboración propia.

Las tareas que se llevaron a cabo en este proceso son:

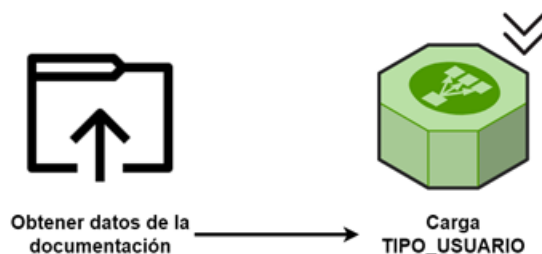
- Inicio: Inicia la ejecución de los procesos en el momento que se indique.
- Carga Subdimensión TIPO\_USUARIO: ejecuta el contenedor de pasos programados en *MySQL Workbench* para cargar la tabla de subdimensión TIPO\_USUARIO.
- Carga Subdimensión UNIDAD\_ACADEMICA: ejecuta el contenedor de pasos programados en *MySQL Workbench* para cargar la tabla de subdimensión UNIDAD\_ACADEMICA.
- Carga Subdimensión BIBLIOTECA\_ADSCRIPCION: ejecuta el contenedor de pasos programados en *MySQL Workbench* para cargar la tabla de subdimensión BIBLIOTECA\_ADSCRIPCION.



- Carga Subdimensión CARRERA: ejecuta el contenedor de pasos programados en *MySQL Workbench* para cargar la tabla de subdimensión CARRERA.
- Carga Subdimensión TEMA\_MDOC: ejecuta el contenedor de pasos programados en *MySQL Workbench* para cargar la tabla de subdimensión TEMA\_MDOC.
- Carga Dimensión USUARIO: ejecuta el contenedor de pasos programados en *MySQL Workbench* para cargar la tabla de dimensión USUARIO.
- Carga Dimensión MDOC: ejecuta el contenedor de pasos programados en *MySQL Workbench* para cargar la tabla de dimensión MDOC.
- Carga Dimensión TIPO\_DEVOLUCION: ejecuta el contenedor de pasos programados en *MySQL Workbench* para cargar la tabla de dimensión TIPO\_DEVOLUCION.
- Carga Dimensión TIPO\_TRANSACCION: ejecuta el contenedor de pasos programados en *MySQL Workbench* para cargar la tabla de dimensión TIPO\_TRANSACCION.
- Carga Dimensión MULTA: ejecuta el contenedor de pasos programados en *MySQL Workbench* para cargar la tabla de dimensión MULTA.
- Carga Dimensión BIBLIOTECARIO: ejecuta el contenedor de pasos programados en *MySQL Workbench* para cargar la tabla de dimensión BIBLIOTECARIO.
- Carga Dimensión TIEMPO: ejecuta el contenedor de pasos programados en *MySQL Workbench* para cargar la tabla de dimensión TIEMPO.
- Carga tabla de hechos CIRCULACION\_MDOC: ejecuta el contenedor de pasos programados en *MySQL Workbench* para cargar la tabla de tabla de hechos CIRCULACION\_MDOC.

## CARGA DE LA SUBDIMENSIÓN TIPO\_USUARIO

La Figura 25 ilustra las tareas llevadas a cabo para la carga de la subdimensión TIPO\_USUARIO.



**Figura 25. Carga de la subdimensión TIPO\_USUARIO**

Fuente: Elaboración propia.

Este proceso es un contenedor de pasos que incluyó las siguientes tareas:

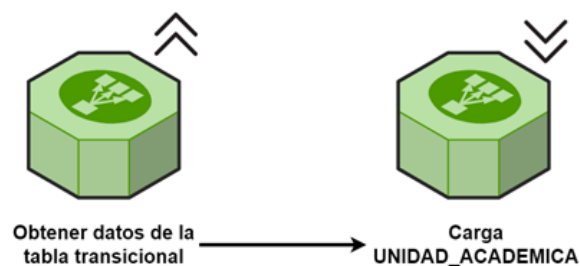
1. Obtener datos de la documentación: Entre las fuentes de información proporcionada se encuentra el documento del catálogo de códigos para los valores que aparecen en las tablas Oracle del Aleph, en los cuales se especifican los tipos de usuarios existentes. Se extrajo esta información para ser ingresada en la subdimensión correspondiente.
2. Carga TIPO\_USUARIO: Para realizar la carga de los datos anteriormente extraídos se realizó la inserción mediante una sentencia *SQL*, en primera instancia se indicó en que tabla serían insertados los datos, TIPO\_USUARIO, continuando con los valores en donde serían insertados, *key\_tipo\_usuario* y *tipo\_usuario*. Fue importante tomar en cuenta el orden de estos elementos, pues fueron ingresados en este formato. Cada valor fue encapsulado mediante paréntesis, además de ser separados por comas.

La sentencia *SQL* configurada para este paso fue la siguiente:

- **INSERT INTO**  
TIPO\_USUARIO (key\_tipo\_usuario,tipo\_usuario)  
**VALUES**  
('1', 'Alumno'),  
('2', 'Académico'),  
('3', 'Administrativo'),  
('4', 'Externo'),  
('5', 'Préstamo'),  
('6', 'Posgrado'),  
('7', 'Sistema Abierto');

## CARGA DE LA SUBDIMENSIÓN UNIDAD\_ACADEMICA

La Figura 26 ilustra las tareas llevadas a cabo para la carga de la subdimensión UNIDAD\_ACADEMICA.



**Figura 26. Carga de la subdimensión UNIDAD\_ACADEMICA**

Fuente: Elaboración propia.

Este proceso es un contenedor de pasos que incluyó las siguientes tareas:

- Obtener datos de la tabla transicional: Se obtuvieron los datos necesarios a partir de una consulta SQL del esquema transicional para cargar la dimensión UNIDAD\_ACADEMICA.
- Se tomó como fuente de entrada la tabla AsignarFacultad del esquema transicional antes mencionado.

- Se obtuvo una lista de identificadores que se relacionan a las unidades académicas disponibles en el sistema. Estos datos fueron obtenidos respectivamente de las columnas idUnidad y Unidad\_Academica de la tabla AsignarFacultad. Esta lista presentó valores únicos en el campo del identificador, mismos que fueron utilizados como llaves primarias en la subdimensión UNIDAD\_ACADEMICA. Se procedió a obtener los datos agrupando estos valores únicos para cumplir con el objetivo antes mencionado.

A continuación, se expone la sentencia SQL configurada para este paso:

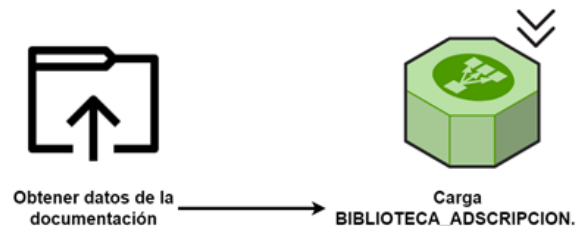
```

• SELECT
    idUnidad, Unidad_Academica
FROM
    AsignarFacultad
GROUP BY idUnidad;

```

## CARGA DE LA SUBDIMENSIÓN BIBLIOTECA\_ADSCRIPCION

La Figura 27 ilustra las tareas llevadas a cabo para la carga de la subdimensión BIBLIOTECA\_ADSCRIPCION.



**Figura 27. Carga de la subdimensión BIBLIOTECA\_ADSCRIPCION**

Fuente: Elaboración propia.

Este proceso es un contenedor de pasos que incluyó las siguientes tareas:

- Obtener datos de la documentación: Se empleó el documento de los códigos de los valores en los que se hace referencia a las bibliotecas de adscripción

existentes. Se extrajo esta información para ser ingresada en la subdimensión correspondiente.

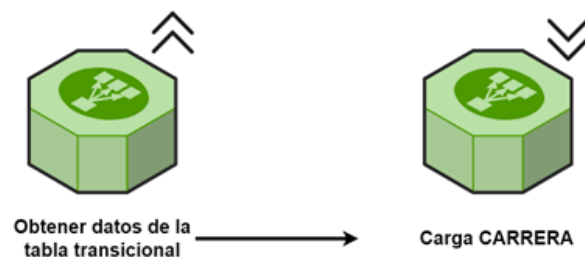
- Carga BIBLIOTECA\_ADSCRIPCION: Para realizar la carga de los datos anteriormente extraídos se realizó la inserción mediante una sentencia *SQL*. En primera instancia se indicó en qué tabla serían insertados los datos BIBLIOTECA\_ADSCRIPCION, continuando con los valores en donde serían insertados, *key\_biblioteca\_adscripcion* y *biblioteca\_adscripcion*. Fue importante tomar en cuenta el orden de estos elementos, pues fueron ingresados en ese formato. Cada valor fue encapsulado mediante paréntesis y fueron separados por comas.

La sentencia *SQL* configurada para este paso fue la siguiente:

```
• INSERT INTO
  BIBLIOTECA_ADSCRIPCION (key_biblioteca_adscripcion,biblioteca_adscripcion)
VALUES
  ('DS', 'DES de Salud');
```

## CARGA DE LA SUBDIMENSIÓN CARRERA

La Figura 28 ilustra las tareas llevadas a cabo para la carga de la subdimensión CARRERA.



**Figura 28. Carga de la subdimensión CARRERA**

Fuente: Elaboración propia.

Este proceso es un contenedor de pasos que incluyó las siguientes tareas:

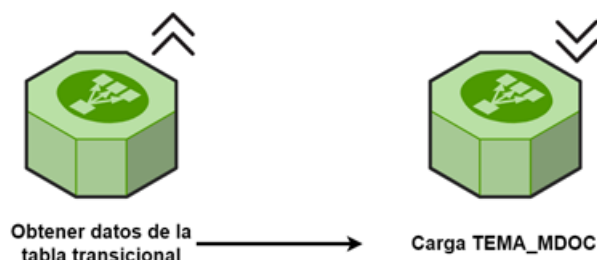
- Obtener datos de la tabla transicional: Se obtuvieron los datos necesarios a partir de una consulta *SQL* del esquema transicional para cargar la dimensión CARRERA.
- Se tomó como fuente de entrada la tabla *AsignarFacultad* del esquema transicional antes mencionado.
- De tal tabla se obtuvo un identificador relacionado a cada una de las carreras disponibles, al nivel académico y al programa educativo al que pertenece el usuario. Estos datos fueron obtenidos de las columnas *CLAVE\_FACULTAD*, *DESC\_PROGRAMA\_EDUCATIVO*, *DESC\_NIVEL\_ACADEMICO* y *PROGRAMA\_EDUCATIVO* respectivamente. Esta lista manejó valores únicos en el campo del identificador que fueron empleados en la llave primaria de la subdimensión CARRERA, por lo cual se procedió a obtener los datos agrupando este valor para cumplir con el objetivo antes mencionado. La sentencia *SQL* configurada en este paso fue la siguiente:

- ```
SELECT
    CLAVE_FACULTAD,
    DESC_PROGRAMA_EDUCATIVO,
    DESC_NIVEL_ACADEMICO,
    PROGRAMA_EDUCATIVO
FROM
    AsignarFacultad
GROUP BY CLAVE_FACULTAD;
```

- Carga CARRERA: Almacena en la tabla de subdimensión CARRERA los datos obtenidos en el paso anterior.

## CARGA DE LA SUBDIMENSIÓN TEMA\_MDOC

La Figura 29 ilustra las tareas llevadas a cabo para la carga de la subdimensión TEMA\_MDOC.



**Figura 29. Carga de la subdimensión TEMA\_MDOC**

Fuente: Elaboración propia.

Este proceso es un contenedor de pasos que incluyó las siguientes tareas:

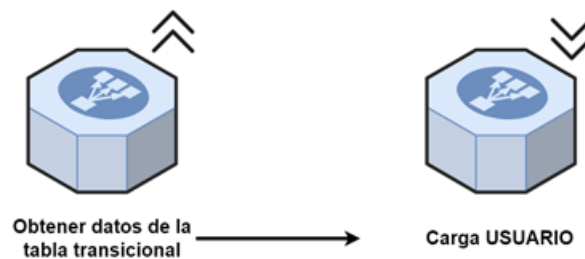
- Obtener datos de la tabla transicional: Se obtuvieron los datos necesarios a partir de una consulta SQL del esquema transicional para cargar la dimensión TEMA\_MDOC.
- Se tomó como fuente de entrada la tabla clasificacion\_temas del esquema transicional antes mencionado.
- De la tabla anterior se obtuvo una lista de identificadores que se relacionan a la clase principal, a la división o subclase y a la sección o grupo temático al que pertenece el MDoc. Estos datos fueron obtenidos de las columnas EXTRACTO, MAIN\_CLASES, DIVISIONS\_SUBCLASES y SECTIONS\_GROUPS. respectivamente. La lista manejó valores únicos en el campo del identificador al ser ingresada en una llave primaria en la subdimensión TEMA\_MDOC, por lo cual se procedió a obtener los datos agrupando este valor para cumplir con el objetivo antes mencionado. La sentencia SQL configurada en este paso fue la siguiente:

- **SELECT**  
EXTRACTO,  
MAIN\_CLASES,  
DIVISIONS\_SUBCLASES,  
SECTIONS\_GROUPS  
**FROM**  
casificacion\_temas  
**GROUP BY** EXTRACTO;

- Carga TEMA\_MDOC: Almacena en la tabla subdimensión TEMA\_MDOC los datos obtenidos en el paso anterior.

## CARGA DE LA DIMENSIÓN USUARIO

La Figura 30 ilustra las tareas llevadas a cabo para la carga de la dimensión USUARIO.



**Figura 30. Carga de la dimensión USUARIO**

Fuente: Elaboración propia.

Este proceso es un contenedor de pasos que incluyó las siguientes tareas:

- Obtener datos de la tabla transicional: Se obtuvieron los datos necesarios a partir de una consulta SQL del esquema transicional para cargar la dimensión USUARIO.
- Se tomó como fuentes de entrada las tablas z303\_z305\_z308 y z305 del esquema transicional antes mencionado.



- Se obtuvo una lista con los identificadores únicos de cada usuario relacionados a la carrera que cursa, su sexo, la clave a la unidad académica a la que pertenece, en qué biblioteca se encuentra adscrito, el tipo de usuario, la fecha de creación de su registro, así como su fecha de actualización y vencimiento. Estos datos fueron obtenidos de las columnas `SUBSTR(Z303\_REC\_KEY,1,12)`, Z303\_FIELD\_3, Z303\_GENDER, SUBSTR(Z303\_FIELD\_3, 1, 4), Z303\_HOME\_LIBRARY, Z305\_BOR\_STATUS, Z303\_OPEN\_DATE, Z303\_UPDATE\_DATE, Z305\_EXPIRY\_DATE respectivamente. Esta lista contiene valores únicos en el campo del identificador que fueron empleados en una llave primaria en la subdimensión USUARIO. Es importante resaltar que en el caso específico de la fecha de vencimiento, ésta se encuentra únicamente en la tabla z305, mientras que el resto de los datos se encontraban en la tabla z303\_z305\_z308. Este problema fue solucionado realizando la unión de las dos tablas aprovechando la llave `SUBSTR(Z303\_REC\_KEY,1,12)` para la tabla z303\_z305\_z308 y `SUBSTR(Z305\_REC\_KEY,1,12)` para la tabla z305. Las llaves primarias deben de ser iguales para tener la unión correcta. Ya unidas se procedió a obtener los datos agrupando este valor para cumplir con el objetivo antes mencionado.

La sentencia SQL configurada en este paso fue la siguiente:

```

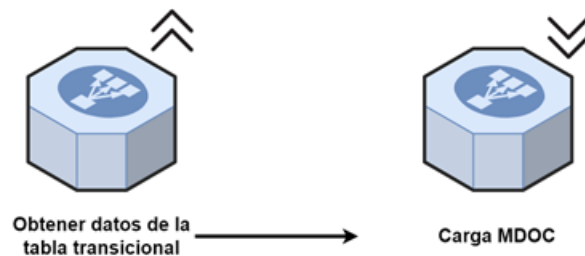
SELECT
  T1.`SUBSTR(Z303_REC_KEY,1,12)`,
  Z303_FIELD_3,
  Z303_GENDER,
  SUBSTR(Z303_FIELD_3, 1, 4),
  Z303_HOME_LIBRARY,
  T1.Z305_BOR_STATUS,
  Z303_OPEN_DATE,
  Z303_UPDATE_DATE,
  Z305_EXPIRY_DATE
FROM
  DWH_TRANS.z303_z305_z308 T1
  LEFT JOIN
  Z305 T2 ON T1.`SUBSTR(Z303_REC_KEY,1,12)` = T2.`SUBSTR(Z305_REC_KEY,1,12)`
GROUP BY Z305_BOR_STATUS;

```

- Carga USUARIO: Almacena en la tabla de dimensión USUARIO los datos obtenidos en el paso anterior.

## CARGA DE LA DIMENSIÓN MDOC

La Figura 31 ilustra las tareas llevadas a cabo para la carga de la dimensión MDOC.



**Figura 31. Carga de la dimensión MDOC**

Fuente: Elaboración propia.

Este proceso es un contenedor de pasos que incluyó las siguientes tareas:

- Obtener datos de la tabla transicional: Se obtuvieron los datos necesarios a partir de una consulta *SQL* del esquema transicional para cargar la dimensión MDOC.
- Se tomó como fuentes de entrada las tablas z303\_z305\_z308 y z305 del esquema transicional antes mencionado.
- Se obtuvo una lista de identificadores únicos que permiten establecer relación con los elementos de título, autor, pie de imprenta, año, clasificación, colección, tipo de material, tipo de adquisición, fecha de ingreso a sistema, fecha de última devolución, estado, cambios, último cambio, la fecha y hora de baja de los MDocs de la BDESSalud, además de la agrupación a que se le asigna por sistema a cada título. Estos datos fueron obtenidos de las columnas Z30H\_BARCODE, Z13\_TITLE, Z13\_AUTHOR, Z13\_IMPRINT, Z13\_YEAR, SUBSTR(Z30H\_CALL\_NO,1), Z30H\_CALL\_NO, Z30H\_COLLECTION, Z30H\_MATERIAL, Z30H\_ITEM\_STATISTIC, Z30H\_OPEN\_DATE, Z30H\_DATE\_LAST\_RETURN, Z30H\_ITEM\_STATUS, Z30H\_H\_REASON,

Z30H\_H\_REASON\_TYPE, Z30H\_H\_DATE, Z30H\_H\_HOUR y Z13\_REC\_KEY, respectivamente. Además, para calcular la llave foránea referente al tema es necesario obtener los primeros dígitos de la clasificación en la columna Z30H\_CALL\_NO para asignarle el tema correspondiente utilizando la tabla clasificacion\_temas.

- Esta lista debió tener valores únicos en el campo del identificador para ser ingresada en una llave primaria en la subdimensión MDOC. En este apartado es necesario indicar que se identificaron datos que era necesario extraer desde la tabla Z13 y el resto de los datos estaban en la tabla Z30H. Este problema fue solucionado al unir las dos tablas aprovechando la llave Z30H\_BARCODE para la tabla Z13 y Z30\_BARCODE para la tabla Z13. Estas llaves debían ser iguales para unir correctamente ambas tablas. Ya unidas se procedió a obtener los datos agrupando este valor para cumplir con el objetivo antes mencionado. Para los casos de MDoc en donde se registró más de un cambio, se eligió el que tuvo la fecha más próxima al momento en que fueron extraídos los datos, para de esta manera asegurar que fuese el último cambio el que quedara registrado en el DW.

La sentencia SQL configurada en este paso fue la siguiente:

- **SELECT**

```

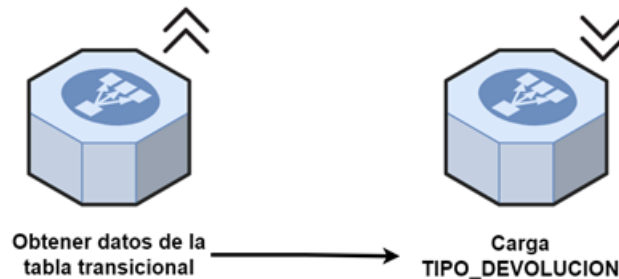
T1.Z30H_BARCODE,
Z13_TITLE,
Z13_AUTHOR,
Z13_IMPRINT,
Z30H_CALL_NO,
Z13_YEAR,
SUBSTR(Z30H_CALL_NO,
      1,
      POSITION(' ' IN Z30H_CALL_NO)) - 1,
Z30H_COLLECTION,
Z30H_MATERIAL,
Z30H_ITEM_STATISTIC
Z30H_OPEN_DATE,
Z30H_DATE_LAST_RETURN,
Z30H_ITEM_STATUS,
Z30H_H_REASON,
Z30H_H_REASON_TYPE,
Z30H_H_DATE,
Z30H_H_HOUR,
Z13_REC_KEY
FROM
  Z30H T1
    INNER JOIN
  Z13 T2 ON T1.Z30H_BARCODE = T2.Z30_BARCODE
GROUP BY Z30H_BARCODE
ORDER BY Z30H_H_DATE DESC;

```

Carga MDOC: Almacena en la tabla de dimensión MDOC los datos obtenidos en el paso anterior.

## CARGA DE LA DIMENSIÓN TIPO\_DEVOLUCION

La Figura 32 ilustra las tareas llevadas a cabo para la carga de la dimensión TIPO\_DEVOLUCION.



**Figura 32. Carga de la dimensión TIPO\_DEVOLUCION**

Fuente: Elaboración propia.

Este proceso es un contenedor de pasos que incluyó las siguientes tareas:

- Obtener datos de la tabla transicional: Se obtuvieron los datos necesarios a través de una consulta *SQL* del esquema transicional para cargar la dimensión TIPO\_DEVOLUCION.
- Se tomó como fuente de entrada la tabla Z36H del esquema transicional antes mencionado.
- Se obtuvo una lista de identificadores, los cuales se relacionan a la fecha de préstamo, fecha de vencimiento, fecha de devolución, hora de préstamo, hora de vencimiento, hora de devolución y número de renovaciones. Estos datos fueron obtenidos de las columnas Z36H\_NUMBER, Z36H\_LOAN\_DATE, Z36H\_DUE\_DATE, Z36H\_RETURNED\_DATE, Z36H\_LOAN\_HOUR, Z36H\_DUE\_HOUR, Z36H\_RETURNED\_HOUR, Z36H\_NO\_RENEWAL. Esta lista presentó valores únicos en el campo del identificador al ser ingresada en una llave primaria en la subdimensión TIPO\_DEVOLUCION. Se procedió a obtener los datos agrupando este valor para cumplir con el objetivo antes mencionado.

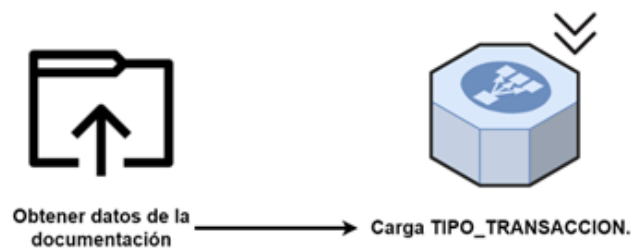
La sentencia SQL configurada en este paso fue la siguiente:

- **SELECT**  
Z36H\_NUMBER,  
Z36H\_LOAN\_DATE,  
Z36H\_DUE\_DATE,  
Z36H\_RETURNED\_DATE,  
Z36H\_LOAN\_HOUR,  
Z36H\_DUE\_HOUR,  
Z36H\_RETURNED\_HOUR,  
Z36H\_NO\_RENEWAL  
**FROM**  
Z36H;

Carga TIPO\_DEVOLUCION: Almacena en la tabla de dimensión TIPO\_DEVOLUCION los datos obtenidos en el paso anterior.

### CARGA DE LA DIMENSIÓN TIPO\_TRANSACCION

La Figura 33 ilustra las tareas llevadas a cabo para la carga de la dimensión TIPO\_TRANSACCION.



**Figura 33. Carga de la dimensión TIPO\_TRANSACCION**

Fuente: Elaboración propia.

Este proceso es un contenedor de pasos que incluyó las siguientes tareas:

- Obtener datos de la documentación: Entre las fuentes de información proporcionadas se encuentra el documento del catálogo de códigos de los valores

empleados en las tablas Oracle Aleph, en el cual se listan los códigos de los tipos transacción existentes. Se extrajo esta información para ser ingresada en la dimensión correspondiente.

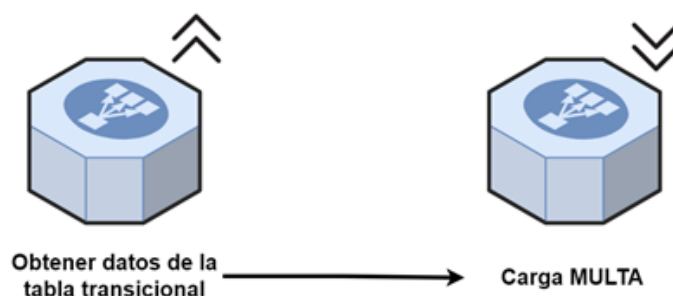
- Carga TIPO\_TRANSACCION: Para realizar la carga de los datos anteriormente extraídos se realizó la inserción mediante una sentencia SQL. En primera instancia se indicó en qué tabla serían insertados los datos, TIPO\_TRANSACCION. Se siguió con la indicación de en dónde serían insertados los valores (key\_tipo\_transaccion, d\_tipo\_transaccion y grupo\_transaccion). Fue importante tomar en cuenta el orden de éstos, debido a que en este formato tuvieron que ser ingresados. La asignación de grupo\_transaccion se realizó de forma arbitraria para agrupar a los tipos de transacción en las categorías de préstamos, renovaciones o devoluciones (si es que pertenecían a alguno de estos grupos). En caso contrario se le asignó el valor de NULL al registro correspondiente. Cada valor fue encapsulado usando paréntesis, además de ser separado por comas.

La sentencia SQL configurada en este paso fue la siguiente:

- ```
INSERT INTO TIPO_TRANSACCION(key_tipo_transaccion, d_tipo_transaccion, grupo_transaccion) VALUES
('50', 'Préstamo externo', 'LOAN'),
('52', 'Circulación fuera de línea', NULL),
('58', 'Préstamo eliminado', NULL),
('61', 'Devolución', 'RETURN'),
('62', 'Renovación - Usuario (Web)', 'RENEW'),
('63', 'Renovación - Bibliotecario', 'RENEW'),
('65', 'MDoc declarado perdido', NULL),
('66', 'MDoc reclamado devuelto', NULL),
('71', 'Reservación', NULL),
('73', 'Solicitud de reserva', NULL),
('74', 'Retener transacción de solicitud (colocada en el MDoc no disponible, sin expansión)', NULL),
('80', 'Préstamo interno', 'LOAN'),
('82', 'Devolución de MDoc no prestado (Error)', 'RETURN'),
('100', 'Transacción tabla z36h', NULL);
```

## CARGA DE LA DIMENSIÓN MULTA

La Figura 34 ilustra las tareas llevadas a cabo para la carga de la dimensión MULTA.



**Figura 34. Carga de la dimensión MULTA**

Fuente: Elaboración propia.

Este proceso es un contenedor de pasos que incluyó las siguientes tareas:

- Obtener datos de la tabla transicional: Se obtuvieron los datos necesarios a partir de una consulta SQL del esquema transicional para cargar la dimensión MULTA.
- Se tomó como fuente de entrada la tabla z31 del esquema transicional antes mencionado.
- Se obtuvo una lista de identificadores que se relacionan al estado de la multa, el bibliotecario que atendió la transacción de multa, el tipo de multa y la fecha y hora del pago. Estos datos fueron obtenidos de las columnas Z31\_REC\_KEY, Z31\_STATUS, Z31\_PAYMENT\_CATALOGER, Z31\_TYPE y Z31\_PAYMENT\_DATE\_KEY, respectivamente.
- Esta lista manejó valores únicos en el campo del identificador al ser ingresada en una llave primaria en la dimensión MULTA. Se procedió a obtener los datos agrupando este valor para cumplir con el objetivo antes mencionado.



La sentencia SQL configurada en este paso fue la siguiente:

```
• SELECT
    Z31_REC_KEY,
    Z31_STATUS,
    Z31_PAYMENT_CATALOGER,
    Z31_TYPE,
    Z31_PAYMENT_DATE_KEY
FROM
    z31
GROUP BY Z31_REC_KEY;
```

Carga MULTA: Almacena en la tabla de dimensión MULTA los datos obtenidos en el paso anterior.

## CARGA DE LA DIMENSIÓN BIBLIOTECARIO

La Figura 35 ilustra las tareas llevadas a cabo para la carga de la dimensión BIBLIOTECARIO.



**Figura 35. Carga de la dimensión BIBLIOTECARIO**

Fuente: Elaboración propia.

Este proceso es un contenedor de pasos que incluyó las siguientes tareas:

- Obtener datos de la tabla transicional: Se obtuvieron los datos necesarios a partir de una consulta SQL del esquema transicional para cargar la dimensión BIBLIOTECARIO.

- Se tomaron como fuentes de entrada las tablas Z35 y Z36H del esquema transicional antes mencionado.
- Se obtuvo una lista de los identificadores de los bibliotecarios. Estos datos fueron obtenidos de las columnas Z35\_CATALOGER\_NAME, Z36H\_LOAN\_CATALOGER\_NAME, Z36H\_RENEW\_CATALOGER\_NAME, Z36H\_RETURN\_CATALOGER\_NAME. En todas ellas se encuentran los identificadores de los bibliotecarios, por lo tanto, fue necesario realizar una unión entre las tablas antes mencionadas. Esta lista manejó valores únicos en el campo del identificador al ser ingresada en una llave primaria en la dimensión BIBLIOTECARIO. Se procedió a obtener los datos agrupando este valor para cumplir con el objetivo antes mencionado.

La sentencia SQL configurada en este paso fue la siguiente:

```

SELECT
    Z35_CATALOGER_NAME
FROM
    Z35
GROUP BY Z35_CATALOGER_NAME
UNION SELECT
    Z36H_LOAN_CATALOGER_NAME
FROM
    Z36H
UNION SELECT
    Z36H_RETURN_CATALOGER_NAME
FROM
    Z36H
UNION SELECT
    Z36H_RENEW_CATALOGER_NAME
FROM
    Z36H;

```

- Cálculo para determinar número de errores de devolución en la dimensión BIBLIOTECARIO: Para obtener el número de errores cometidos referente a la columna errordevolucion\_bibliotecario, fue necesario contabilizar las

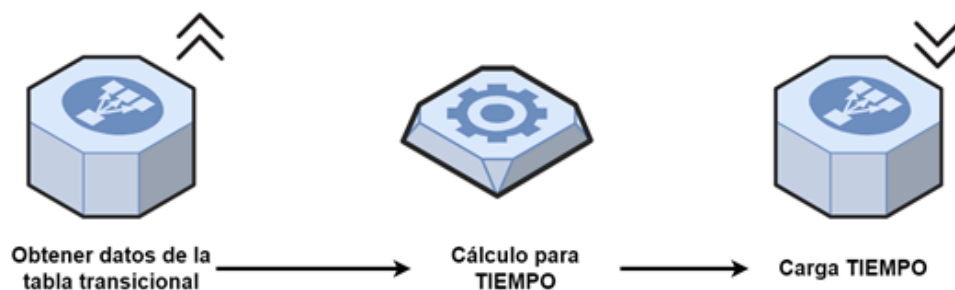
transacciones referentes al tipo de transacción con valor 82 'Devolución de MDOC no prestado (Error)' de la columna Z35\_EVENT\_TYPE y relacionarlo con el identificador de bibliotecario que realizó esta transacción (campo Z35\_CATALOGER\_NAME), ambos ubicados en la tabla Z35. En los casos en que no se encontraron errores relacionados al identificador de bibliotecario, a éste se le asignó el valor de 0 en el campo referido a errordevolucion\_bibliotecario. Posteriormente se configuró la siguiente sentencia SQL para llevar a cabo este paso:

- **SELECT**  
    Z35\_CATALOGER\_NAME, COUNT(\*)  
**FROM**  
    Z35\_2  
**WHERE**  
    Z35\_EVENT\_TYPE = 82  
**GROUP BY** Z35\_CATALOGER\_NAME;

Carga BIBLIOTECARIO: Almacena en la tabla de dimensión BIBLIOTECARIO los datos obtenidos en los pasos anteriores.

## CARGA DE LA DIMENSIÓN TIEMPO

La Figura 36 ilustra las tareas llevadas a cabo para la carga de la dimensión TIEMPO.



**Figura 36. Carga de la dimensión TIEMPO**

Fuente: Elaboración propia.

Este proceso es un contenedor de pasos que incluyó las siguientes tareas:

- Obtener datos de la tabla transicional: Se obtuvieron los datos necesarios a partir de una consulta *SQL* del esquema transicional para cargar la dimensión TIEMPO.
- Se tomó como fuentes de entrada las tablas Z35 y Z36H del esquema transicional antes mencionado.
- Se obtuvo una lista de un identificador del tiempo, este dato fue obtenido de las columnas Z35\_EVENT\_DATE, Z35\_EVENT\_HOUR y Z36H\_TIME. En todas ellas se encuentran los identificadores del tiempo de las transacciones, por lo tanto, fue necesario unir las tablas antes mencionadas.
- Para Z35\_EVENT\_DATE y Z35\_EVENT\_HOUR es necesario realizar una concatenación de estos datos para obtener el formato requerido para el identificador, en el caso específico en que la hora no sea mayor a 100000 (el suceso sucedió antes de las 10 horas) se incumplirá el formato necesario por lo cual es necesario concatenarle un '0' antes de la hora para que se cumpla esta restricción. Esta lista debe de tener valores únicos en el campo del identificador al ser ingresada en una llave primaria en la dimensión TIEMPO, se procede a obtener los datos agrupando este valor para cumplir con el objetivo antes mencionado.

La sentencia *SQL* configurada en este paso fue la siguiente:

```

SELECT
    SUBSTR(Z36H_TIME,1,14)
FROM
    Z36H
GROUP BY SUBSTR(Z36H_TIME,1,14)
UNION
SELECT
    concat(Z35_EVENT_DATE, Z35_EVENT_HOUR)
FROM
    Z35_2
WHERE Z35_EVENT_HOUR >= 100000
UNION
SELECT
    concat(Z35_EVENT_DATE,'0', Z35_EVENT_HOUR)
FROM
    Z35_2
WHERE Z35_EVENT_HOUR < 100000

```

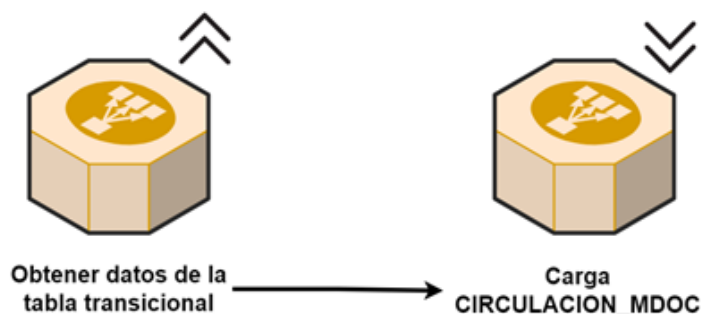
Cálculo para dimensión TIEMPO: Para obtener los datos referentes a fecha, hora, mes\_num, mes\_text, anio, semestre, semana y dia\_habil se extrajo la información de las tablas fuente cuando fue posible o, en caso contrario, se ingresaron funciones para calcular el dato correspondiente. Para ello se configuró la sentencia SQL que se muestra a continuación:

- `UPDATE TIEMPO`  
`SET`  
`mes_num = MONTH(key_tiempo),`  
`anio = YEAR(key_tiempo),`  
`semana = WEEK(key_tiempo),`  
`dia_habil = DAYOFWEEK(key_tiempo),`  
`fecha = DATE(key_tiempo),`  
`hora = TIME(key_tiempo);`
- `update TIEMPO set mes_text='ENERO' where mes_num=1;`
- `update TIEMPO set mes_text='FEBRERO' where mes_num=2;`
- `update TIEMPO set mes_text='MARZO' where mes_num=3;`
- `update TIEMPO set mes_text='ABRIL' where mes_num=4;`
- `update TIEMPO set mes_text='MAYO' where mes_num=5;`
- `update TIEMPO set mes_text='JUNIO' where mes_num=6;`
- `update TIEMPO set mes_text='JULIO' where mes_num=7;`
- `update TIEMPO set mes_text='AGOSTO' where mes_num=8;`
- `update TIEMPO set mes_text='SEPTIEMBRE' where mes_num=9;`
- `update TIEMPO set mes_text='OCTUBRE' where mes_num=10;`
- `update TIEMPO set mes_text='NOVIEMBRE' where mes_num=11;`
- `update TIEMPO set mes_text='DICIEMBRE' where mes_num=12;`
- `update TIEMPO set semestre=1 where mes_num<=6;`
- `update TIEMPO set semestre=2 where mes_num>6;`
- `update TIEMPO set dia_habil='Do' where dia_habil='1';`
- `update TIEMPO set dia_habil='Lu' where dia_habil='2';`
- `update TIEMPO set dia_habil='Ma' where dia_habil='3';`
- `update TIEMPO set dia_habil='Mi' where dia_habil='4';`
- `update TIEMPO set dia_habil='Ju' where dia_habil='5';`
- `update TIEMPO set dia_habil='Vi' where dia_habil='6';`
- `update TIEMPO set dia_habil='Sa' where dia_habil='7';`

Carga TIEMPO: Almacena en la tabla de dimensión TIEMPO los datos obtenidos en los pasos anteriores.

## CARGA DE LA TABLA DE HECHOS CIRCULACION\_MDOC

La Figura 37 ilustra las tareas llevadas a cabo para la carga de la tabla de hechos CIRCULACION\_MDOC.



**Figura 37. Carga de la tabla de hechos CIRCULACION\_MDOC**

Fuente: Elaboración propia.

Este proceso es un contenedor de pasos que incluyó las siguientes tareas:

- Obtener datos de la tabla transicional: Se obtuvieron los datos necesarios a partir de una consulta SQL del esquema transicional para cargar la dimensión CIRCULACION\_MDOC.
- Se tomó como fuentes de entrada las tablas Z35 y Z36H del esquema transicional antes mencionado.
- Se obtuvo una lista de identificadores, el cual se relaciona entre sí para dar como resultado a una transacción única. Estos identificadores hacen referencia al tiempo en que sucedió la transacción, tipo de transacción, el MDoc involucrado, el usuario, el bibliotecario, la multa (en caso de que exista), y el tipo de devolución (en el caso que exista). Estos datos fueron extraídos de: Z303\_REC\_KEY, Z35\_EVENT\_DATE, Z35\_EVENT\_HOUR, Z35\_EVENT\_TYPE, Z30\_BARCODE, Z35\_ID, Z35\_CATALOGER\_NAME para el caso de la tabla Z35, mientras que para la tabla Z36H se obtuvo esta información de las columnas Z36H\_NUMBER, Z36H\_TIME, Z30\_BARCODE, Z36H\_ID, Z36H\_LOAN\_CATALOGER\_NAME, Z31\_REC\_KEY y Z36H\_NUMBER.

- En el caso de la tabla Z35, para obtener la fecha en el formato requerido fue necesario concatenar el contenido de las columnas Z35\_EVENT\_DATE AS DATE, Z35\_EVENT\_HOUR, además de agregar los caracteres '00' al final para cumplir con el formato establecido.
- De misma forma para la tabla Z36H fue necesario obtener una subcadena de 14 caracteres de la columna Z36H\_TIME, además de ingresar el valor de '100' en el valor de TIPO\_TRANSACCION para poder identificar su procedencia.
- Esta lista debe de tener valores únicos en los campos de identificadores al ser ingresada en las llaves primarias en la tabla de hechos CIRCULACION\_MDOC.

A continuación, se exponen las sentencias SQL configuradas para este paso:

- **SELECT DISTINCT**  
 Z36H\_NUMBER,  
 CAST(SUBSTR(Z36H\_TIME, 1, 14) AS DATETIME),  
 '100',  
 T1.Z30\_BARCODE,  
 Z36H\_ID,  
 Z36H\_LOAN\_CATALOGER\_NAME,  
 Z31\_REC\_KEY,  
 Z36H\_NUMBER  
 FROM  
 DWH\_TRANS.Z36H T1  
 LEFT JOIN  
 aux\_multa ON Z31\_NUMBER = Z36H\_NUMBER  
 GROUP BY CAST(SUBSTR(Z36H\_TIME, 1, 14) AS DATETIME) , '100' , T1.Z30\_BARCODE , Z36H\_ID , Z36H\_LOAN\_CATALOGER\_NAME;
- **SELECT**  
 Z303\_REC\_KEY,  
 CONCAT(CAST(Z35\_EVENT\_DATE AS DATE),  
 ' ',  
 CAST(Z35\_EVENT\_HOUR \* 100 AS TIME)),  
 Z35\_EVENT\_TYPE,  
 Z30\_BARCODE,  
 Z35\_ID,  
 Z35\_CATALOGER\_NAME  
 FROM  
 DWH\_TRANS.Z35  
 GROUP BY CONCAT(CAST(Z35\_EVENT\_DATE AS DATE),  
 ' ',  
 CAST(Z35\_EVENT\_HOUR \* 100 AS TIME)) , Z35\_EVENT\_TYPE , Z30\_BARCODE , Z35\_ID , Z35\_CATALOGER\_NAME;



Carga CIRCULACION\_MDOC: Almacena en la tabla de hechos CIRCULACION\_MDOC los datos obtenidos en el paso anterior.

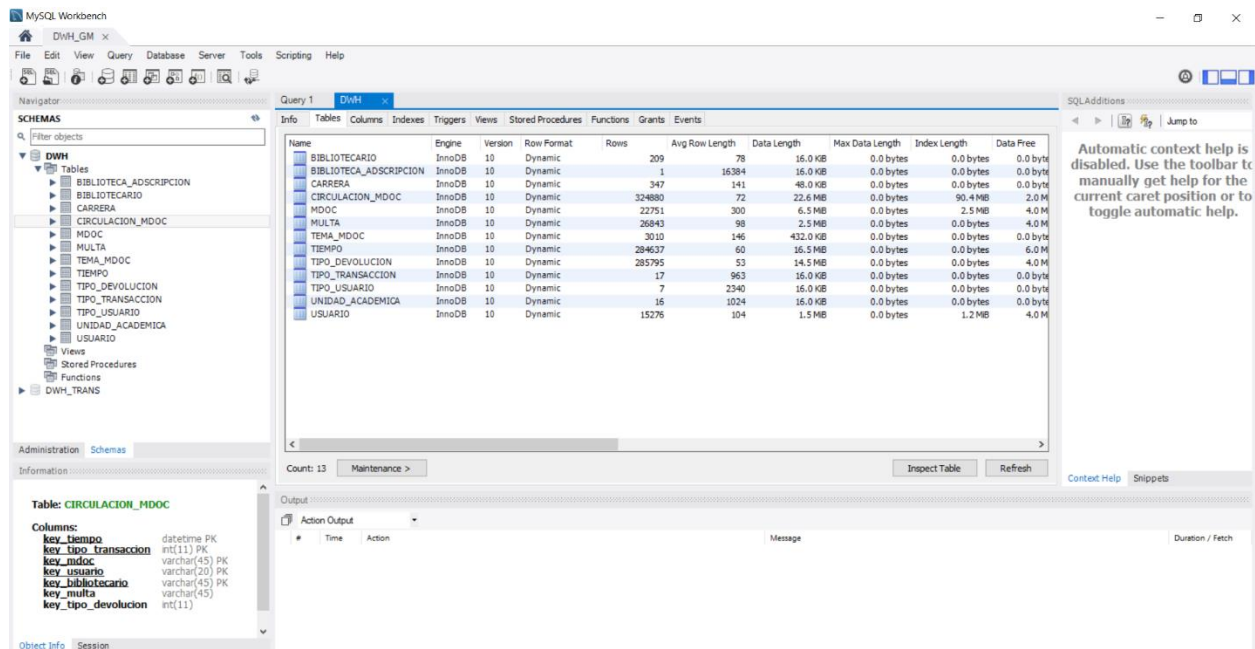
### **2.2.4.3 Cubo multidimensional**

Luego de la carga de los datos y siguiendo la metodología Hefesto, se empleó *MySQL Workbench* para fue crear un cubo multidimensional configurado a partir del modelo lógico indicado con anterioridad:

1. Se creó un cubo multidimensional llamado DW BDESSalud.
2. Se indicó que la tabla de hechos es CIRCULACION\_MDOC.
3. Se crearon las siguientes dimensiones y subdimensiones:
  - Dimensión MDOC
  - Subdimensión TEMA\_MDOC
  - Dimensión USUARIO
  - Subdimensión TIPO\_USUARIO
  - Subdimensión CARRERA
  - Subdimensión UNIDAD\_ACADEMICA
  - Subdimensión BIBLIOTECA\_ADSCRIPCION
  - Dimensión TIPO\_TRANSACCION
  - Dimensión TIPO\_DEVOLUCION
  - Dimensión BIBLIOTECARIO
  - Dimensión MULTA
  - Dimensión TIEMPO
4. Se creó el indicador llamado NÚMERO DE PRÉSTAMOS, el cual fue calculado empleando la función de agregación COUNT.
5. Se creó el indicador llamado NÚMERO DE RENOVACIONES, el cual fue calculado empleando la función de agregación COUNT.
6. Se creó el indicador llamado NÚMERO DE DEVOLUCIONES, el cual fue calculado empleando la función de agregación COUNT.

7. Se creó el indicador llamado NÚMERO DE ERRORES DE DEVOLUCIÓN DE MDOC, el cual fue calculado empleando la función de agregación COUNT.
8. Se creó el indicador llamado NÚMERO DE MULTAS, el cual fue calculado empleando la función de agregación COUNT.
9. Se creó el indicador llamado NÚMERO DE MULTAS COBRADAS, el cual fue calculado empleando la función de agregación COUNT.
10. Se creó el indicador llamado NÚMERO DE MULTAS CANCELADAS, el cual fue calculado empleando la función de agregación COUNT.
11. Se creó el indicador llamado NÚMERO DE DEVOLUCIONES PREVIAS, el cual fue calculado empleando la función de agregación COUNT.
12. Se creó el indicador llamado NÚMERO DE DEVOLUCIONES EN TIEMPO, el cual fue calculado empleando la función de agregación COUNT.
13. Se creó el indicador llamado NÚMERO DE DEVOLUCIONES VENCIDAS, el cual fue calculado empleando la función de agregación COUNT.

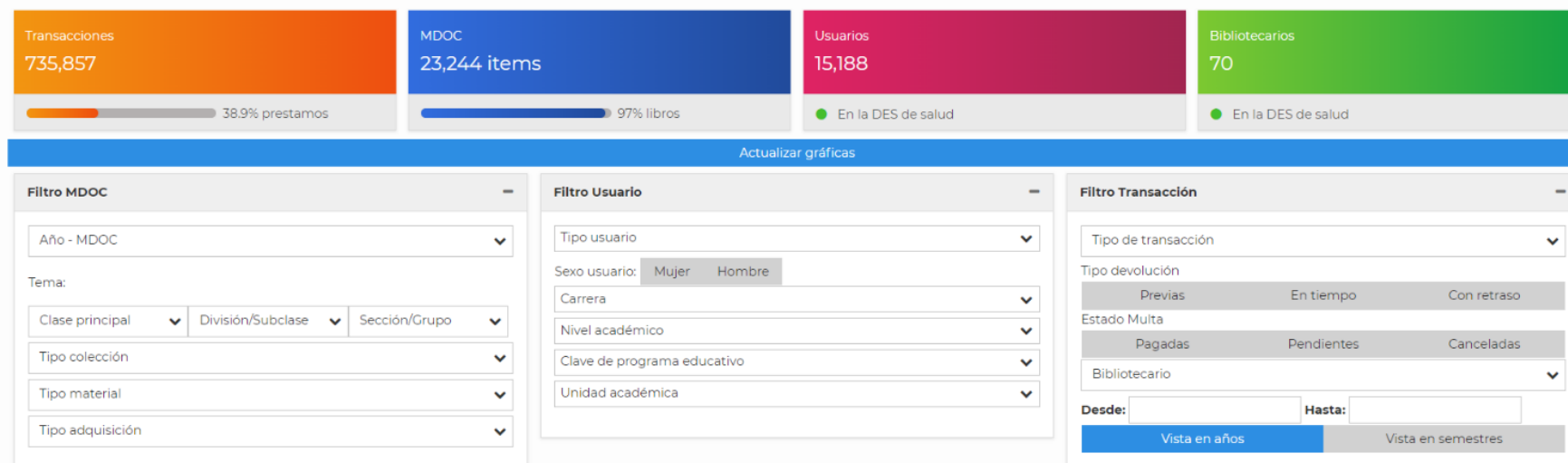
La Figura 38 muestra los componentes del cubo DW BDESSalud.



**Figura 38. Componentes del Cubo DW BDESSalud en MySQL Workbench**

Fuente: Elaboración propia.

El medio para la consulta y visualización de los datos contenidos en el DW fue una aplicación web dinámica desarrollada en el lenguaje de programación Java que consulta y consume los datos del DW a partir de consultas en lenguaje *Java Script*. La capa de presentación para la interfaz con la cual interactúan los usuarios finales se creó empleando componentes para diseño de interfaces de usuario de *Prime Faces*. La Figura 39 muestra la pantalla principal del prototipo de interfaz de consulta y visualización de datos del DW de la BDESSalud.



**Figura 39. Prototipo de interfaz de consulta y visualización del DW de la BDESSalud**

Fuente: Página web de la Interfaz de consulta y visualización del DW de la BDESSalud.

## **2.3 Diseño y aplicación de la bibliominería**

### **2.3.1 Nota previa: bibliominería y la privacidad de los usuarios**

De acuerdo con Fayyad, Piatetsky-Shapiro y Smyth (1996), al trabajar con bases de datos que contienen información personal, se debe tener cuidado de abordar adecuadamente los asuntos éticos y legales sobre la invasión a la privacidad. Los autores Cox y Jantti (2012) recomiendan que, si [una] biblioteca desea crear su propio [proyecto de análisis de datos mediante MD y DW], deberá asegurarse de poder hacerlo conforme a las leyes de privacidad y protección de datos pertinentes (p. 310).

Es importante considerar que, desde hace tiempo, muchas personas utilizan productos y servicios a sabiendas de que recopilan datos o incluso requieren el intercambio de grandes cantidades de información personal (por ejemplo, Facebook o dispositivos que dan seguimiento a ciertas actividades). Estos individuos han decidido hacerlo luego de determinar que los beneficios de utilizar tales tecnologías bien valen la pena de compartir ciertos datos. Éstos pueden ser utilizados por algoritmos informáticos para determinar qué información se mostrará a usuarios específicos (llámense resultados de búsqueda, recomendaciones, contenido sugerido, etc.) creando diferentes experiencias fundamentadas en lo que la tecnología determina que los usuarios desean o necesitan ver. Las bibliotecas, como organizaciones que recopilan datos y están interesadas en mejorar sus productos y servicios, pueden encontrar oportunidades de utilizar los datos para sus propios fines o se les puede solicitar que compartan sus datos con empresas, gobiernos u otras organizaciones. Ambas propuestas plantean desafíos éticos relacionados con el usuario, su privacidad y libertad intelectual (American Library Association, 2014).

Desde hace una década, las tendencias detectadas por las principales organizaciones o asociaciones bibliotecarias en el mundo en nuestra área de conocimiento incluyen de manera recurrente el tema de las cuestiones éticas relacionadas al manejo de los datos de usuarios y las implicaciones que estas actividades tienen en la preservación de su

privacidad<sup>54</sup>. Inclusive la Federación Internacional de Asociaciones e Instituciones Bibliotecarias (en adelante IFLA, por sus siglas en inglés) emitió en 2013 una declaración sobre minería de datos y de texto, en la cual indicaba que tales actividades son herramientas esenciales para el avance del aprendizaje y nuevas formas de creación, sin embargo su preocupación iba por el lado de cómo las regulaciones sobre derecho de autor podrían afectar su implementación, más que por el de la protección de la privacidad de los usuarios.

En 2015, la IFLA compartió su declaratoria sobre la privacidad en el entorno bibliotecario he hizo hincapié en la responsabilidad de los profesionales de la información de buscar los medios para proteger lo mejor posible la privacidad y los datos personales de los usuarios ante la “recopilación excesiva de datos” que supone el uso de ciertos servicios informativos. De acuerdo con la IFLA “[l]os servicios de biblioteca e información pueden decidir qué tipo de datos personales recopilarán sobre los usuarios y considerar los principios de seguridad, gestión, almacenamiento, intercambio y retención de datos” (IFLA, 2015, p. 2). Luego, en el 2020, la IFLA emitió una declaratoria sobre las bibliotecas y la inteligencia artificial (IA) (IFLA, 2020), en la cual estableció lo que denomina principios clave que deben regir la aplicación de IA y aprendizaje automático (ML) en los centros de información y en los cuales se encuentra, nuevamente, el tema de la privacidad de los datos de los usuarios. La IFLA señala que la aplicación de estos desarrollos debe ponerse al servicio de la innovación y bienestar públicos y que, con la preparación necesaria (considerando las preocupaciones éticas y limitaciones actuales), las bibliotecas pueden utilizar de manera responsable estas tecnologías para avanzar en su misión social.

Un denominador común a la realización de los estudios relacionados con la aplicación de minería de datos en bibliotecas tiene que ver con la preservación de los datos originados por las transacciones realizadas en estos centros de información. Es aconsejable que los administradores bibliotecarios contemplen el impacto a largo plazo que puede tener el deshacerse de los registros, pues si se tomara la previsión de

---

<sup>54</sup> Véanse: Association of College and Research Libraries, Research Planning and Review Committee (2018 y 2020).

salvaguardar la información de interés en almacenes de datos (DW), se podrían llevar a cabo análisis de importancia para la toma de decisiones sin dejar de proteger la confidencialidad y privacidad de sus usuarios (Nicholson, 2003a). Conforme se reúne información más detallada, las bases de datos se tornan más útiles, empero, se vuelve más difícil la tarea de preservar la confidencialidad (Nicholson, 2006). Nicholson indicó el uso de sustitutos demográficos<sup>55</sup>, los cuales solucionan este problema y le brindan al usuario la certeza de que sus datos no serán empleados de manera indebida.

Cómo universidad autónoma estatal, la UACH se rige por la Ley de Protección de Datos Personales del Estado de Chihuahua, razón por la cual el SUBA cuenta con una Política de seguridad de datos<sup>56</sup>. En cuanto a los datos que pueden considerarse como sensibles, aquellos que fueron empleados para la configuración el DW de la BDESSalud incluyeron únicamente el ID de usuario (ya sea matrícula o número de empleado). Su uso fue estrictamente necesario, pues tal dato constituye una importante llave primaria que permitió enlazar a los usuarios con las transacciones realizadas dentro del proceso de circulación del MDoc analizado en el modelo de DW. Puede considerarse que los datos permanecen seguros y bajo el resguardo de la BDESSalud, pues la interfaz de consulta y visualización de los datos contenidos en el DW requiere de un nombre de usuario y contraseña para su acceso y consulta, misma que únicamente fue proporcionada a la jefa de Servicios Bibliotecarios de la BDESSalud. Asimismo, el uso de tal información fue necesaria para realizar un proceso que busca la mejora de los servicios otorgados a la comunidad usuaria de la BDESSalud. En el caso del ejercicio de bibliominería llevado a cabo, no se incluyó en su realización datos que puedan comprometer la seguridad de los usuarios. A continuación, se exponen cuáles fueron las características del ejercicio de bibliominería implementado en este proyecto.

---

<sup>55</sup> Para más información sobre el tema de los sustitutos demográficos véanse Nicholson (2003a), Nicholson (2006d, p. 3) y Nicholson y Arnott (2007).

<sup>56</sup> Véase Anexo 10: Política de seguridad de datos del SUBA-UACH (Sistema Universitario de Bibliotecas Académicas, 2019).

### **2.3.2 Programa para realizar el ejercicio de bibliominería: Python**

El programa que se utilizó para la aplicación del proceso de bibliominería fue Python<sup>57</sup>. Éste es uno de los lenguajes de programación más usados en el ámbito académico y científico<sup>58</sup> y una de sus utilidades es el entrenamiento de algoritmos para minería de datos. Su difusión y extendida adopción son debido a que utiliza un lenguaje multiplataforma<sup>59</sup>, es de código abierto (y por ende gratuito), además de que se puede integrar con otros lenguajes de programación<sup>60</sup>. Se seleccionó este programa porque es de acceso abierto, su uso es generalizado y se espera que el lenguaje y adaptabilidad de Python permita que en un futuro se pueda profundizar más en el desarrollo de análisis complementarios que la BDESSalud requiera. En esta tesis el programa Python se usó para explotar y analizar parte del contenido de la base de datos del DW de la BDESSalud, en específico, los datos relacionados con los patrones de préstamo y patrones de devolución de MDoc. En el siguiente apartado se detalla cuál fue la técnica de minería de datos elegida y cuál fue el objetivo del análisis de bibliominería efectuado.

### **2.3.3 Aplicación de un proceso de bibliominería en la BDESSalud**

En el caso de esta tesis, se utilizó el proceso de bibliominería para agrupar a usuarios de la BDESSalud que cuentan con perfiles similares, los cuales fueron determinados considerando ciertas características inherentes a los usuarios (sexo, tipo de usuario, carrera, unidad académica), así como la información transaccional que denota los patrones de comportamiento informativo de los usuarios con respecto al uso del MDoc (número de préstamos, categorías temáticas más empleadas, características de devolución del MDoc solicitado en préstamo, etc.). El conjunto de datos empleado se

---

<sup>57</sup> Ver: <https://www.python.org/>

<sup>58</sup> Es muy utilizado para aplicaciones web, Ciencia de datos, inteligencia artificial y análisis y automatización de datos.

<sup>59</sup> Es decir, que es compatible para Windows, Linux/Unix, Mac OS.

<sup>60</sup> Se puede integrar con los lenguajes C, C ++, COM, ActiveX, CORBA y Java.



extraído del DW de la BDESSalud. A continuación, se presentan los pormenores del empleo del proceso de bibliominería.

## **DEFINICIÓN DEL PROBLEMA A RESOLVER CON BIBLIOMINERÍA**

¿Qué características (sexo, tipo de usuario, carrera, unidad académica) tienen los usuarios que solicitan MDoc en préstamo en la BDESSalud? ¿Cuáles son sus patrones de comportamiento informativo con respecto al uso del MDoc de la BDESSalud (considerando la información transaccional referida al número de préstamos, categorías temáticas más empleadas, actualidad del MDoc empleado y características de la devolución del MDoc solicitado en préstamo)?

## **OBJETIVO DE LA APLICACIÓN DE BIBLIOMINERÍA**

Creación de **perfiles de usuarios** de acuerdo con sus **características inherentes** y con la **información transaccional** que denota sus **patrones de comportamiento informativo** con respecto al uso de MDoc de la BDESSalud.

## **ELECCIÓN DEL ALGORITMO DE MINERÍA DE DATOS**

Se recurrió a la técnica de agrupamiento con el algoritmo *K-Means* para identificar conjuntos de usuarios que comparten perfiles similares, los cuales fueron delimitados de acuerdo con sus patrones de uso del MDoc de la BDESSalud. El agrupamiento es una tarea descriptiva común en la que se busca identificar a un conjunto finito de categorías o grupos para describir los datos. Las categorías pueden ser mutuamente excluyentes y exhaustivas o consistir en una representación más [completa], como las categorías jerárquicas o superpuestas (Fayyad, Piatetsky-Shapiro y Smyth, 1996, p. 14).

De acuerdo con Xu y Tian (2015) la agrupación, [es] considerada como [uno de los elementos] más importantes del aprendizaje no supervisado, se ocupa de la partición de

la estructura de datos en un área [de conocimiento] desconocida y constituye la base para un aprendizaje posterior. Sin embargo, no hay un acuerdo para [establecer] una definición completa [con respecto a esta técnica de MD], por lo que se pueden considerar los siguientes elementos de una definición clásica:

- Las instancias en un mismo grupo deben ser lo más similares posible,
- Las instancias en los distintos grupos deben ser lo más diferentes posible,
- La medida de similitud y diferencia debe ser clara y tener un significado práctico

El proceso estándar de agrupamiento se puede dividir en los siguientes pasos:

1. Extracción y selección de características más representativas del conjunto de datos original.
2. Diseño del algoritmo de agrupamiento de acuerdo con las características del problema.
3. Evaluación de resultados de la agrupación y revisión de la validez del algoritmo.
4. Explicación práctica del resultado de la agrupación

La distancia (disimilitud) y la similitud [o cercanía] son la base para construir algoritmos de agrupamiento. En cuanto a las características de los datos cuantitativos, se prefiere la distancia para reconocer la relación entre datos. Y se prefiere la similitud cuando se trata de características de datos cualitativos (Xu y Tian, 2015, p. 166. Traducción propia).

De acuerdo con las recomendaciones de Raina y Krishnamurthy, para llevar a cabo la aplicación de la técnica de agrupamiento primero es importante decidir a qué características o atributos [variables] de las observaciones nos enfocaremos. Una vez creados los grupos en función de tales atributos, se analizarán para comprender cuáles son exactamente las similitudes dentro de cada grupo, lo cual da como resultado un nombre y una descripción útil para cada uno de ellos. Finalmente, tales conocimientos pueden ser útiles para determinar [acciones apropiadas para atender las necesidades] de cada grupo de [usuarios] (Raina y Krishnamurthy 2022, p. 76).

## VARIABLES SELECCIONADAS PARA EL PROCESO DE BIBLIOMINERÍA

A partir de las variables inherentes y transaccionales consideradas en la definición del problema para la aplicación de la bibliominería se construyeron variables apropiadas para crear una vista minable que permitió analizar el comportamiento informativo de los usuarios al aplicar y ejecutar el algoritmo seleccionado empleando el programa Python. De manera natural, los usuarios con mayor antigüedad tuvieron un mayor número de transacciones relacionadas al MDoc (préstamos, renovaciones, devoluciones, etc.), por lo cual se calcularon promedios mensuales como métricas comparables con independencia de la antigüedad del usuario. De esta manera, las variables inherentes y transaccionales consideradas para estructurar la matriz de entrenamiento del algoritmo *K-Means* fueron las que se muestran en la Tabla 30.

**Tabla 30. Variables inherentes y transaccionales para la matriz de entrenamiento**

#	Variable	Fuente	Tipo de dato	Descripción
1	Key usuario	DW BDESSalud	Nominal	Sirve para identificar al usuario.
2	Sexo (usuario)	DW BDESSalud	Nominal	Indica el sexo del usuario.
3	Tipo de usuario	DW BDESSalud	Nominal	Indica el tipo de usuario.
4	Carrera	DW BDESSalud	Nominal	Indica la carrera a que pertenece el usuario.
5	Facultad o Unidad académica	DW BDESSalud	Nominal	Indica la Unidad académica a que pertenece el usuario.
6	Promedio mensual préstamos	DW BDESSalud	Continuo	Señala el promedio mensual de préstamos realizados por el usuario.
7	Promedio de préstamos en meses con actividad	DW BDESSalud	Continuo	Muestra el promedio de préstamos en los meses en que el usuario registró actividad en el sistema.
8	Antigüedad promedio del MDoc consultado	DW BDESSalud	Continuo	Muestra cuál fue la antigüedad promedio del MDoc consultado por el usuario.
9	Promedio de días de retraso por préstamo	DW BDESSalud	Continuo	Muestra el promedio del número de días de retraso en la devolución del MDoc.

#	Variable	Fuente	Tipo de dato	Descripción
10	Porcentaje de devoluciones tardías	DW BDESSalud	Continuo	Señala el porcentaje de devoluciones tardías del MDoc solicitado por el usuario.
11	Promedio de temas generales distintos consultados	DW BDESSalud	Continuo	Indica cuál fue el número de temas distintos de los MDoc solicitados en préstamo por el usuario.
12	Temas consultados	DW BDESSalud	Nominal	Indica el nombre de los temas consultados por el usuario.

Fuente: Elaboración propia.

En el caso de la variable 12, se consideró el gran número de temas existentes en la colección y el problema de exceso de dimensionalidad que éste podría causar al emplear el algoritmo *K-Means*, por lo cual se tomó la decisión de indagar cuáles fueron los grupos temáticos con un mayor número de consultas de parte de los usuarios de la biblioteca. Así, se eligieron dieciséis temas que concentraron el 85% de las consultas en los datos analizados. La Tabla 31 muestra los temas seleccionados para el análisis.

**Tabla 31. Temas con la mayor concentración de consultas en la BDESSalud**

#	Tema - Grupo	Usuarios que consultaron el MDoc en el grupo	% de usuarios
1	<i>Human physiology</i>	3,619	75.36
2	<i>Human anatomy, cytology, histology</i>	3,362	70.01
3	<i>Diseases</i>	2,937	61.16
4	<i>Pharmacology and therapeutics</i>	2,279	47.46
5	<i>Biochemistry</i>	1,991	41.46
6	<i>Medicine &amp; health</i>	1,324	27.57
7	<i>Gynecology, obstetrics, pediatrics, geriatrics</i>	1,009	21.01
8	<i>Surgery, regional medicine, dentistry, ophthalmology, otology, audiology</i>	900	18.74
9	<i>Knowledge</i>	794	16.53
10	<i>Personal health and safety</i>	778	16.20
11	<i>Food technology</i>	675	14.06
12	<i>Physiology and related subjects</i>	524	10.91

#	Tema - Grupo	Usuarios que consultaron el MDoc en el grupo	% de usuarios
13	<i>Chemistry and allied sciences</i>	473	9.85
14	<i>Food and drink</i>	456	9.50
15	<i>Forensic medicine; incidence of injuries, wounds, disease; public preventive medicine</i>	453	9.43
16	<i>Occupational ethics</i>	300	6.25

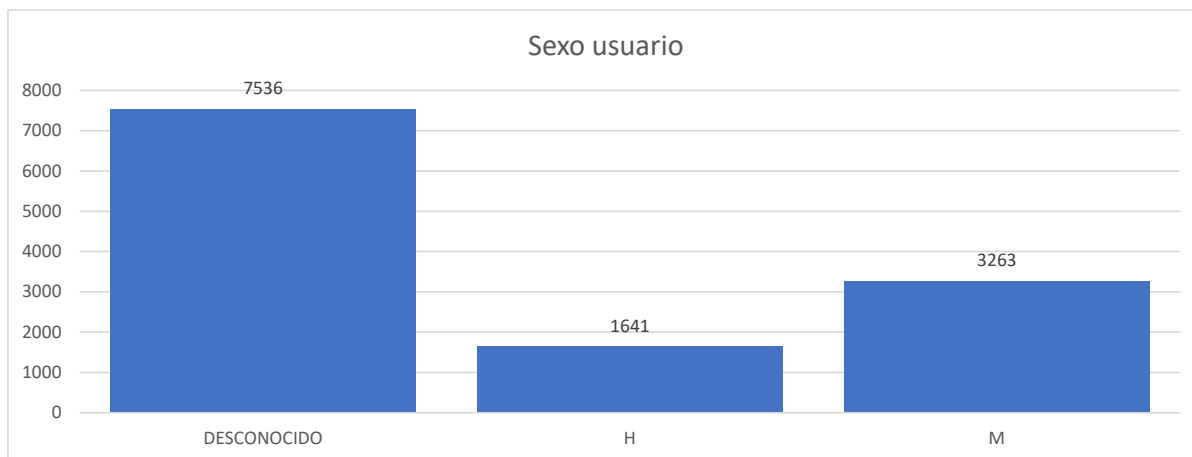
Fuente: Elaboración propia.

## APLICACIÓN DEL PROCESO DE BIBLIOMINERÍA

Una vez identificadas las variables y luego de configurar una consulta *SQL* para obtener los datos necesarios del DW de la BDESSalud para el ejercicio, se llevó a cabo su extracción y se procedió a explorarlos a partir de la implementación de distribuciones estadísticas asociadas a las características inherentes de los usuarios<sup>61</sup> que registraron al menos una transacción en la BDESSalud durante el periodo del 01 de enero de 2010 al 31 de diciembre de 2019. Se contó con información transaccional de 12,441 usuarios. Sin embargo, fue posible identificar que el 60% (7,553) de las ID de usuarios (*key\_usuario*) registradas no contaban con la información inherente necesaria, pues las ID registradas en las transacciones no encontraron correspondencia en la tabla Z303 proporcionada por la Unidad de Sistemas del SUBA, referente a la información de los usuarios de la BDESSalud. Los datos que aparecen con valor desconocido en las figuras subsecuentes tienen lugar debido a que no fueron incluidos en la información del usuario al momento de darlo de alta, ya sea por error u omisión del bibliotecario encargado de realizar este proceso o por errores en la precarga de datos provenientes del SEGA.

En cuanto al sexo del usuario, se encontró que el 60.5% de los registros no cuenta con sexo asignado, como se muestra en la gráfica de la Figura 40.

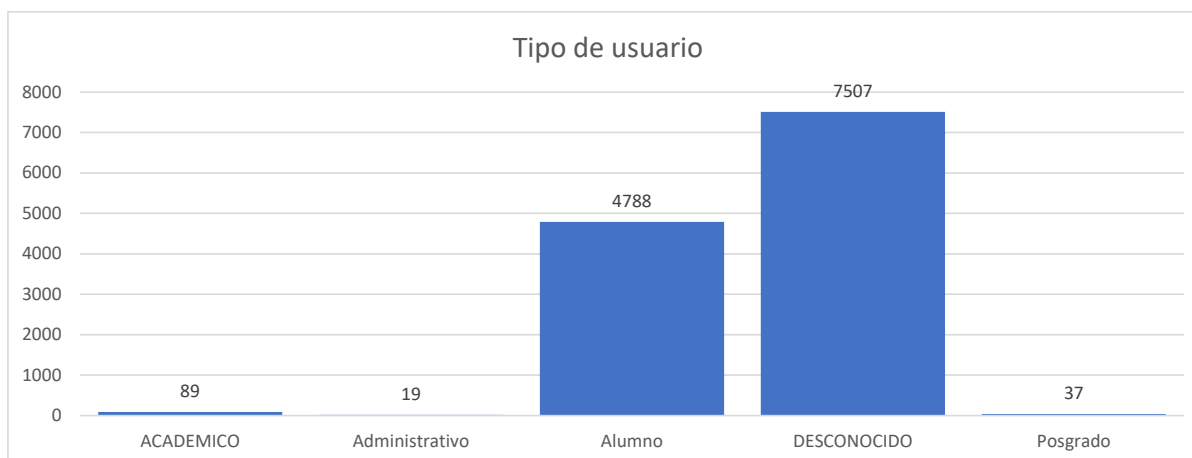
<sup>61</sup> Sexo, tipo de usuario, carrera y unidad académica.



**Figura 40. Sexo del usuario**

Fuente: Elaboración propia.

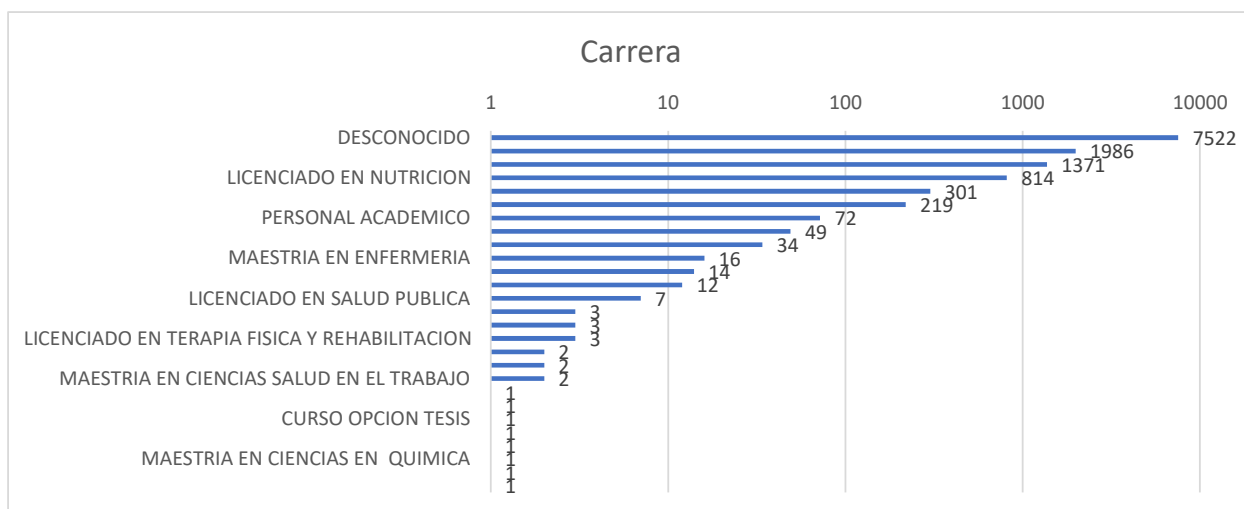
La Figura 41 muestra el apartado de tipo de usuario, en el cual se detectó que no se identifica el tipo del 60.3% usuarios.



**Figura 41. Tipo de usuario**

Fuente: Elaboración propia.

La Figura 42 nos permite verificar que, en lo referente a la carrera, aproximadamente el 60% de los usuarios no la tiene registrada.



**Figura 42. Carrera**

Fuente: Elaboración propia.

La Tabla 32 muestra que, en lo referente a la Unidad Académica o Facultad de procedencia del usuario, se detectó que 7,524 usuarios (60.4%) no tienen Facultad asignada.

**Tabla 32. Frecuencia de usuarios por Facultad**

Facultad	Frecuencia
DESCONOCIDO	7,524
FACULTAD DE ENFERMERIA Y NUTRIOLOGIA	2,537
FACULTAD DE MEDICINA Y CIENCIAS BIOMEDICAS	2,370
FACULTAD DE ODONTOLOGIA	3
FACULTAD DE CIENCIAS DE LA CULTURA FISICA	1
FACULTAD DE CIENCIAS QUIMICAS	1
FACULTAD DE CONTADURIA Y ADMINISTRACION	1
FACULTAD DE FILOSOFIA Y LETRAS	1
FACULTAD DE ZOOTECNIA Y ECOLOGIA	1
UNIDAD CENTRAL DES DE SALUD	1
TOTAL	12,440

Fuente: Elaboración propia.

Es posible que esta falta de correspondencias se deba a la pérdida de información derivada de los procesos de depuración que se realizan en la base de datos de usuarios de la BDESSalud de manera periódica, así como a fallas presentadas en el Aleph debido a problemas en el servidor (reportadas en las actas del CCSUBA), las cuales resultaron en una pérdida de datos que aún no ha sido posible dimensionar de manera certera. Debido a lo anterior, el ejercicio de bibliominería se llevó a cabo considerando el restante 40% (4,888)<sup>62</sup> de los ID de usuarios que sí contaron tanto con la información inherente necesaria para el estudio como con la información de las transacciones realizadas en el sistema. Se descartaron las ID de usuarios que no contaron con información inherente y también se descartaron los registros asociados las facultades de Ciencias de la Cultura Física (1 registro), Ciencias Químicas (1 registro), Contaduría y Administración (1 registro), Filosofía y Letras (1 registro), Zootecnia y Ecología (1 registro) y la Unidad Central DES de Salud (1 registro). Esto debido a que la escasa cantidad de registros puede influir negativamente en la calidad de los grupos de usuarios identificados.

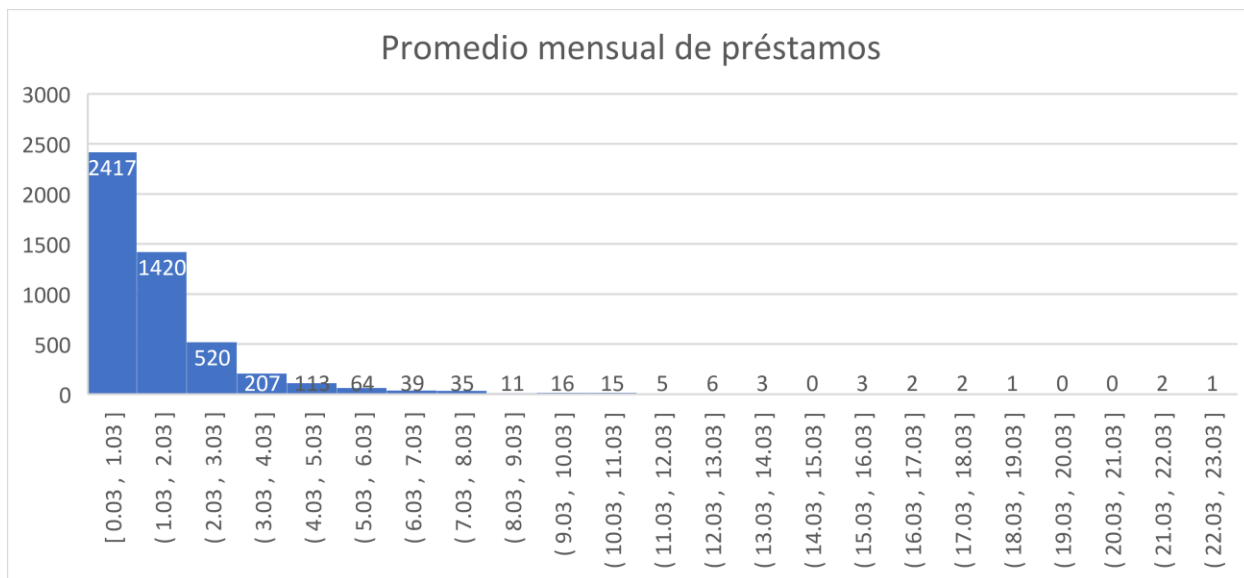
A continuación, se muestra la exploración de las variables transaccionales consideradas en el análisis:

- De acuerdo con el gráfico de la Figura 43, el promedio mensual de préstamos solicitados (considerando el número de meses entre el primer y último préstamo) va de 0 a 3 préstamos para el 93% de los usuarios considerados en el estudio.

---

<sup>62</sup> En la implementación de la matriz de entrenamiento del algoritmo se eliminaron 81 registros más por ausencia de valores, por lo cual se tuvo un set de entrenamiento efectivo de 4,801 registros de usuarios.

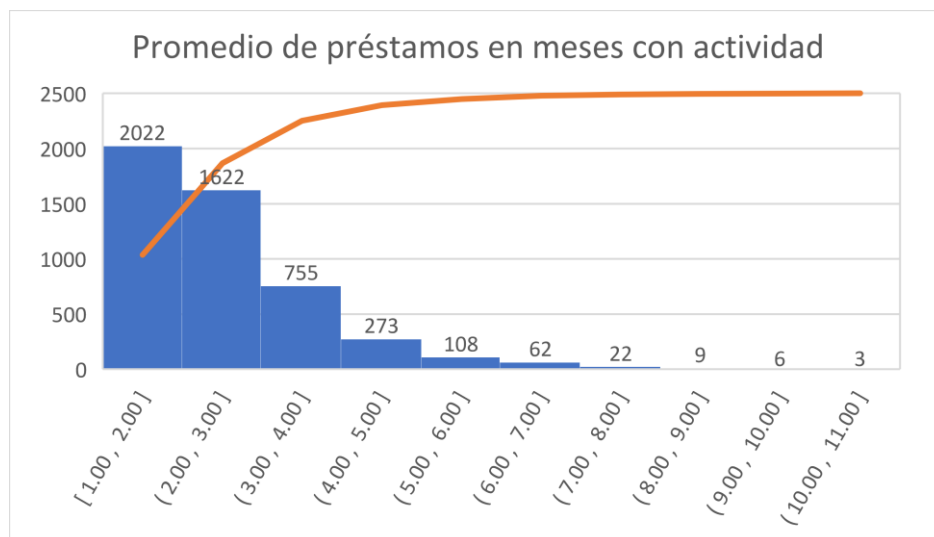




**Figura 43. Promedio mensual de préstamos de MDoc**

FUENTE: Elaboración propia.

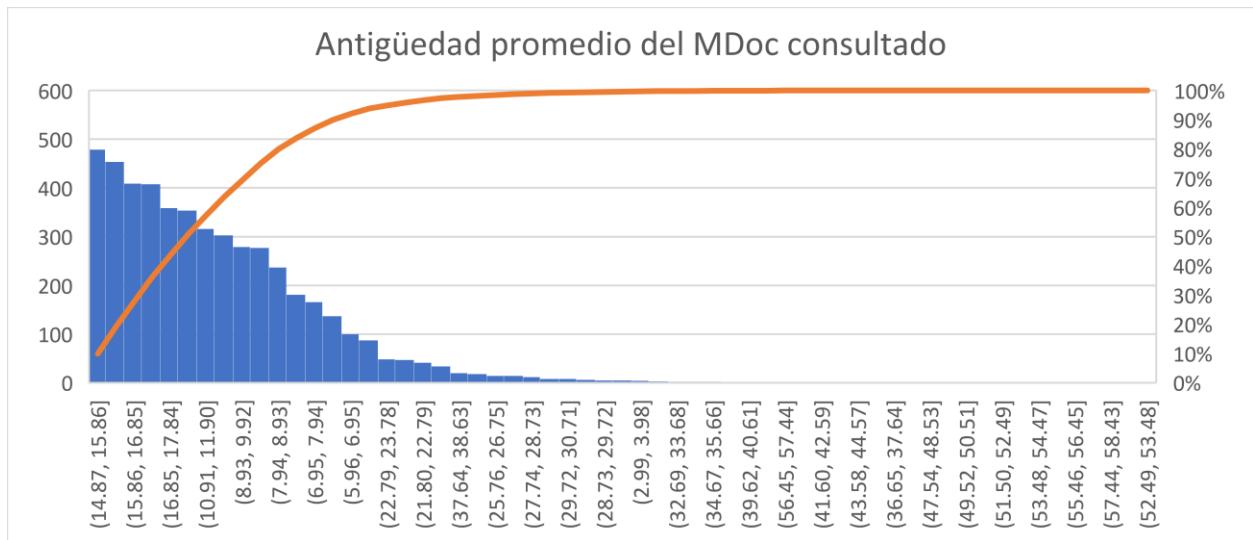
- Promedio mensual de préstamos considerando únicamente los meses con actividad.



**Figura 44. Promedio de préstamos en meses con actividad**

FUENTE: Elaboración propia.

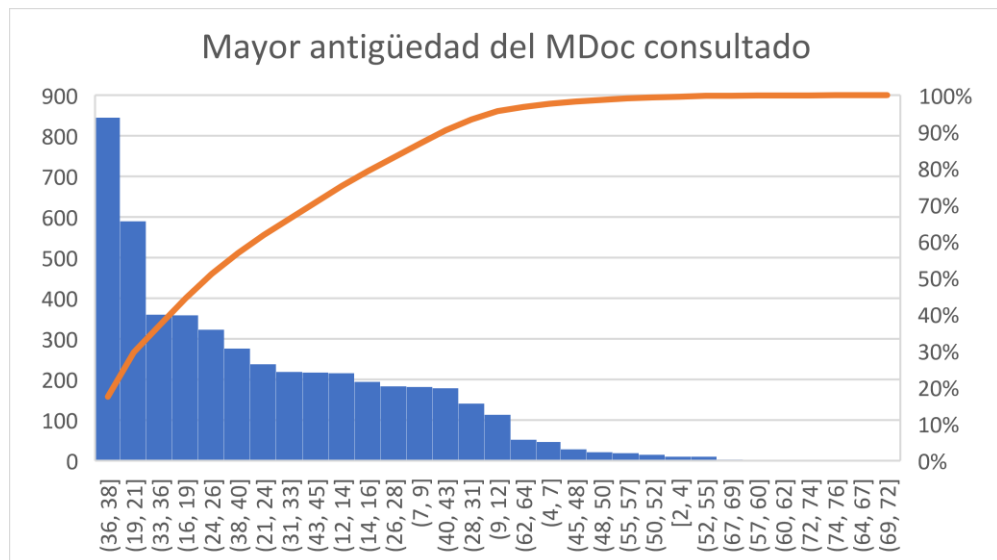
- Antigüedad promedio del MDoc consultado (en años).



**Figura 45. Antigüedad promedio del MDoc consultado**

FUENTE: Elaboración propia.

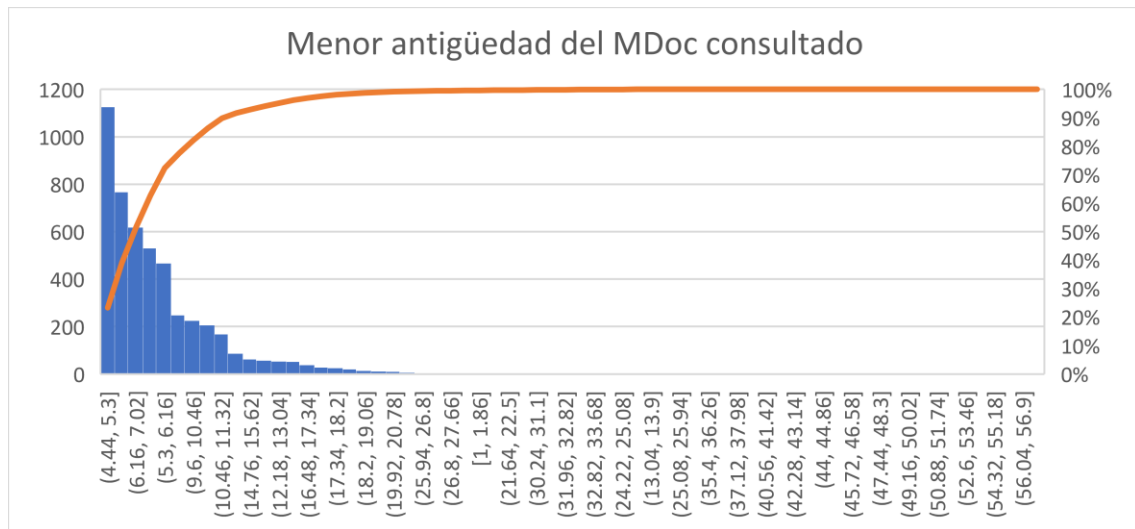
- Mayor antigüedad del MDoc consultado (en años).



**Figura 46. Mayor antigüedad del MDoc consultado**

FUENTE: Elaboración propia.

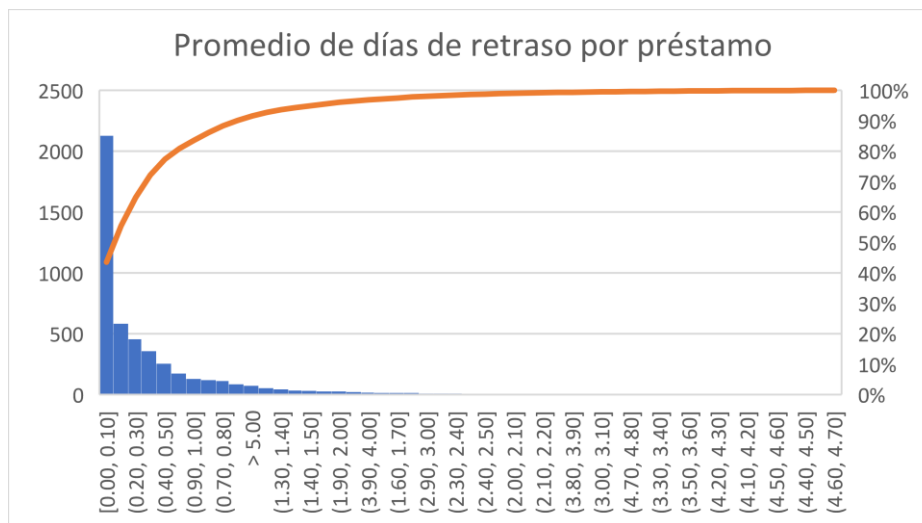
- Menor antigüedad del MDoc consultado (en años).



**Figura 47. Menor antigüedad del MDoc consultado (en años)**

FUENTE: Elaboración propia.

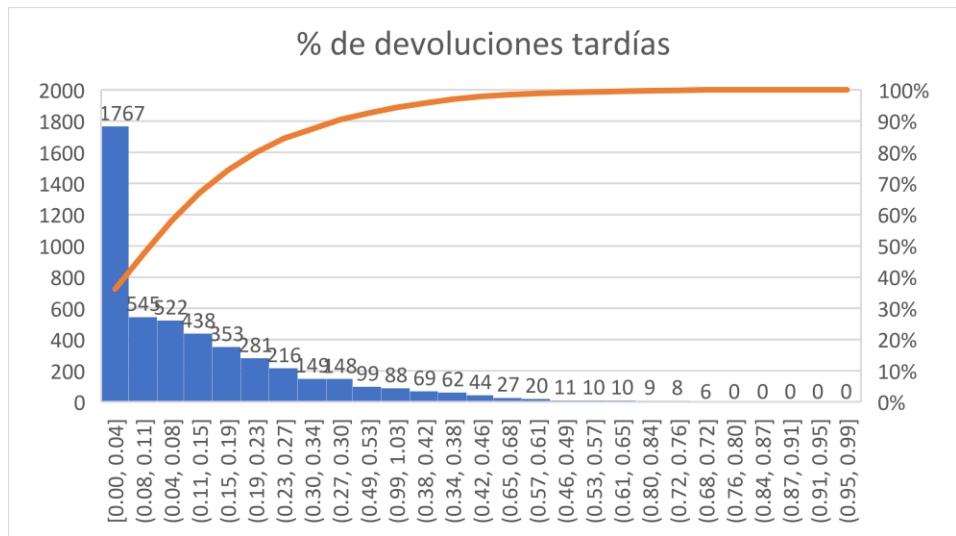
- Promedio de días de retraso por préstamo.



**Figura 48. Promedio de días de retraso por préstamo**

FUENTE: Elaboración propia.

- Porcentaje de devoluciones tardías.



**Figura 49. Porcentaje de devoluciones tardías**

FUENTE: Elaboración propia.

- Número de grupos temáticos consultados.



**Figura 50. Número de grupos temáticos consultados**

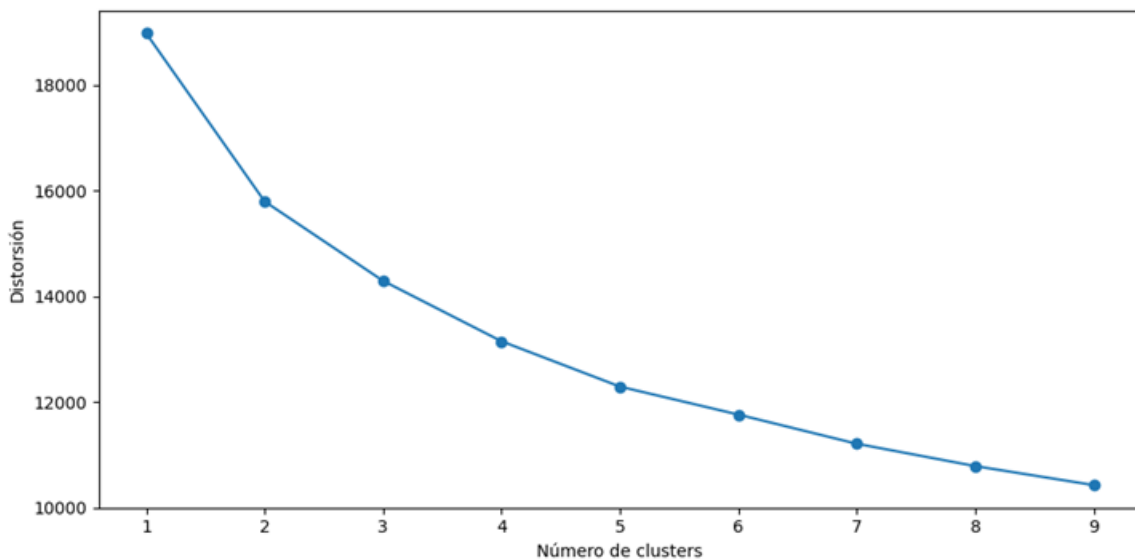
FUENTE: Elaboración propia.

- Temas consultados. Como se mencionó con anterioridad, se seleccionaron los temas que concentraron el mayor número de consultas con el fin de llevar a cabo el análisis. Luego de verificar su frecuencia de uso se obtuvieron dieciséis temas cuyo número, aunque reducido significativamente, también supuso una dimensionalidad alta para la matriz de entrenamiento donde se concentraron las variables. La alternativa tomada para evitar mayores problemas de dimensionalidad que pudieran interferir con un resultado satisfactorio del ordenamiento fue, en principio, la aplicación del método *one-hot encoder*, el cual permitió la modificación de la variable nominal de tema para ser incluida en el análisis. De esta manera se crearon columnas para cada uno de los temas, en las cuales se llevó un marcaje por cada registro donde el valor 1 indica que el usuario empleó el tema en cuestión y el 0 implica que no lo utilizó.

## **ANÁLISIS DE IDENTIFICACIÓN DE GRUPOS O PERFILES EMPLEANDO K-MEANS**

Luego de la revisión de las variables inherentes y transaccionales, se eliminaron los registros con valores nulos: de la matriz de entrenamiento se eliminaron 81 registros por ausencia de valores, lo cual derivó en un set de entrenamiento efectivo de 4,801 registros. Posteriormente y considerando que el algoritmo *K-Means* es sensible a la escala de las variables, se procedió a una normalización de los datos con el método de mínimos y máximos. Uno de los problemas de los algoritmos de clasificación no supervisados (como el *K-Means*), es que requieren como parámetro que se especifiquen el número de agrupaciones ( $k$ ) a generar, lo cual puede impactar en la calidad del agrupamiento. Para asegurar una adecuada aplicación del algoritmo se empleó el método del codo para estimar el número de agrupamientos óptimo, considerando el cálculo de distorsión. Si  $k$  aumenta la distorsión decrementa debido a que las instancias estarán más cercanas a los centroides a los cuales fueron asignadas. La idea fue identificar el valor de  $k$  en donde la distorsión comenzó a aumentar más rápidamente para determinar el número de grupos óptimo.

1. Rango de grupos a probar: de 1 a 9 grupos
2. 10 corridas de *K-Means* para cada grupo
3. Máximo de 300 iteraciones para cada corrida
4. Los resultados se muestran en la Figura 51, que ilustra la gráfica de codo elaborada para determinar el número óptimo de grupos. La gráfica se lee de derecha a izquierda. El primer cambio de pendiente notorio se aprecia en  $c=5$ , a partir de ahí la pendiente es relativamente constante hasta llegar a  $c=2$ , en donde se aprecia el cambio de pendiente más grande. Esto quiere decir que los datos procesados se pueden agrupar eficientemente en dos clases, no obstante, agrupaciones en 3, 4 y 5 clases resultan factibles. En el caso del análisis de bibliominería desarrollado, se eligió clasificar a los usuarios en cuatro grupos con el fin de tener espacios de representatividad e inclusión.

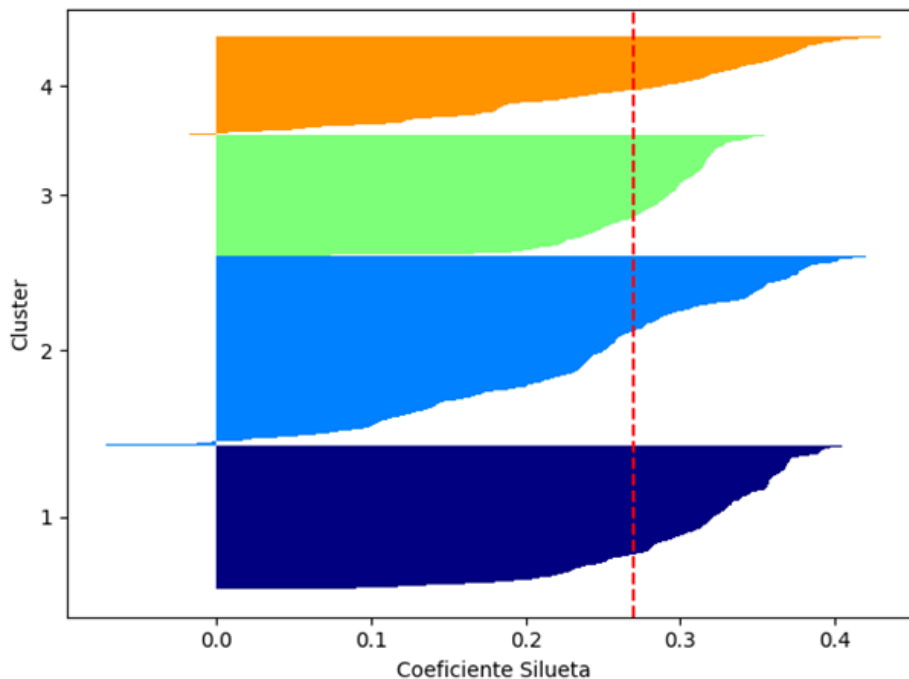


**Figura 51. Gráfica de codo para determinar del número óptimo de grupos**

Fuente: Elaboración propia.

## EVALUACIÓN DE LA APLICACIÓN DEL ALGORITMO DE BIBLIOMINERÍA

La evaluación de la aplicación del algoritmo *K-Means* se llevó a cabo empleando el coeficiente de silueta, una medida intrínseca que consiste en medir qué tan fuertemente están agrupadas las instancias de cada grupo y con base en ello determinar su calidad. El coeficiente de silueta tiene un rango de  $[-1,1]$ . Los valores que representan una mayor calidad de agrupación son los cercanos a 1. Se efectuó el ejercicio para 2, 3, 4 y 5 grupos (factibles de acuerdo con la gráfica de codo anteriormente mostrada). Sin embargo, debido a la alta dimensionalidad de las variables elegidas el resultado no fue del todo satisfactorio (coeficiente de silueta de 0.16 para la agrupación de usuarios en 4 categorías), por lo cual se requirió la aplicación de un análisis de componentes principales (PCA) para reducir la dimensionalidad. Luego de la aplicación del PCA y considerando el escenario de la agrupación de 4 categorías de usuarios, se obtuvo una mejora en el promedio del coeficiente de silueta, al pasar de 0.16 a 0.27, como se muestra en la Figura 52.



**Figura 52. Coeficiente de silueta para 4 grupos de usuarios**

Fuente: Elaboración propia.

# CAPÍTULO 3. Análisis de resultados

Luego de la estructuración del DW de la BDESSalud se continuó con el diseño de un prototipo de interfaz para la consulta y visualización de datos, además de llevar a cabo el ejercicio de bibliominería empleando los datos contenidos en el DW. En este capítulo se mostrarán ejemplos de las consultas que es posible realizar empleando la interfaz creada, así como los resultados del ejercicio de bibliominería.

## 3.1 Resultados de consultas hechas al data warehouse de la BDESSalud

Esta sección muestra ejemplos de las consultas que es posible realizar en el prototipo de interfaz para la consulta y visualización de los datos contenidos en el DW de la BDESSalud. La interfaz consiste en un tablero de control (*dashboard*) que permite a sus usuarios interactuar con los elementos del modelo de DW elaborado en el segundo capítulo de esta tesis<sup>63</sup>. Se consideró el empleo de gráficos interactivos para la representación visual de las consultas planteadas. Para elaborar las consultas y/o explorar los datos, el usuario debe definir qué filtros empleará según los elementos que necesite para realizar el análisis requerido. Los filtros específicos habilitados en la interfaz se muestran en la Tabla 33.

**Tabla 33. Filtros empleados para la consulta de los indicadores**

#	Filtro	Opciones
1	AÑO DEL MDOC	Opción múltiple
2	TEMA - CLASE PRINCIPAL	Opción múltiple
3	TEMA - DIVISIÓN/SUBCLASE	Opción múltiple
4	TEMA - SECCIÓN/GRUPO	Opción múltiple
5	TIPO DE COLECCIÓN	Opción múltiple
6	TIPO DE MATERIAL	Opción múltiple
7	TIPO DE ADQUISICION	Opción múltiple
8	TIPO DE PRÉSTAMO	Opción múltiple

<sup>63</sup> Los elementos que el usuario del tablero de control puede emplear para hacer los reportes son los indicadores principales del modelo: número de préstamos, renovaciones, devoluciones, errores de devolución del MDoc, multas, multas cobradas, multas canceladas, devoluciones tempranas, devoluciones en tiempo y devoluciones vencidas.



#	Filtro	Opciones
9	TIPO DE RENOVACIÓN	Opción múltiple
10	TIPO DE USUARIO	Opción múltiple
11	SEXO DEL USUARIO	Opción múltiple
12	CARRERA	Opción múltiple
13	NIVEL ACADÉMICO	Opción múltiple
14	PROGRAMA EDUCATIVO	Opción múltiple
15	UNIDAD ACADÉMICA	Opción múltiple
16	TIPO DE TRANSACCIÓN	Opción múltiple
17	TIPO DE DEVOLUCIÓN	Opción múltiple
18	ESTADO DE LA MULTA	Opción múltiple
19	BIBLIOTECARIO	Opción múltiple
20	TIEMPO (rango de fechas, año(s)/semestre(s))	Opción múltiple

Fuente: Elaboración propia.

A continuación, se muestran algunos ejemplos de consultas que es posible efectuar empleando el prototipo de interfaz para la consulta y visualización de los datos contenidos en el DW de la BDESSalud:

### **CONSULTA 1. Títulos más consultados agrupados por tema, tipo de usuario y carrera**

**NARRATIVA:** Este reporte muestra los quince títulos más consultados agrupados por tema (desagregado a nivel Sección/Grupo temático), por un tipo de usuario determinado, de una carrera específica, en un determinado tiempo.

**¿QUÉ SE CONSULTA?** Los 15 títulos (MDOC) más consultados agrupados por TEMAS (desagregados a nivel Sección/Grupo) más consultados en la BDESSalud por un TIPO DE USUARIO específico de una CARRERA determinada en un TIEMPO determinado.

**¿CÓMO SE CONFIGURA ESTA CONSULTA?** La exploración de los datos comienza con la determinación de un espacio temporal específico (TIEMPO: rango de fechas, años, semestres), posteriormente se filtra por CARRERA, la gráfica va definida en principio por TIPO DE USUARIO, luego se despliegan los TEMAS (a nivel Sección/Grupo) del MDOC más solicitados en PRÉSTAMO y finalmente se muestran los 15 títulos más consultados (agrupados en el TEMA a que correspondan).

**EJEMPLO DE CONSULTA:** ¿Cuáles fueron los títulos más consultados agrupados por temas (desagregados a nivel Sección/Grupo) más consultados en la BDESSalud por los estudiantes de la Licenciatura en Nutrición en el último semestre (agosto-diciembre) de 2019?

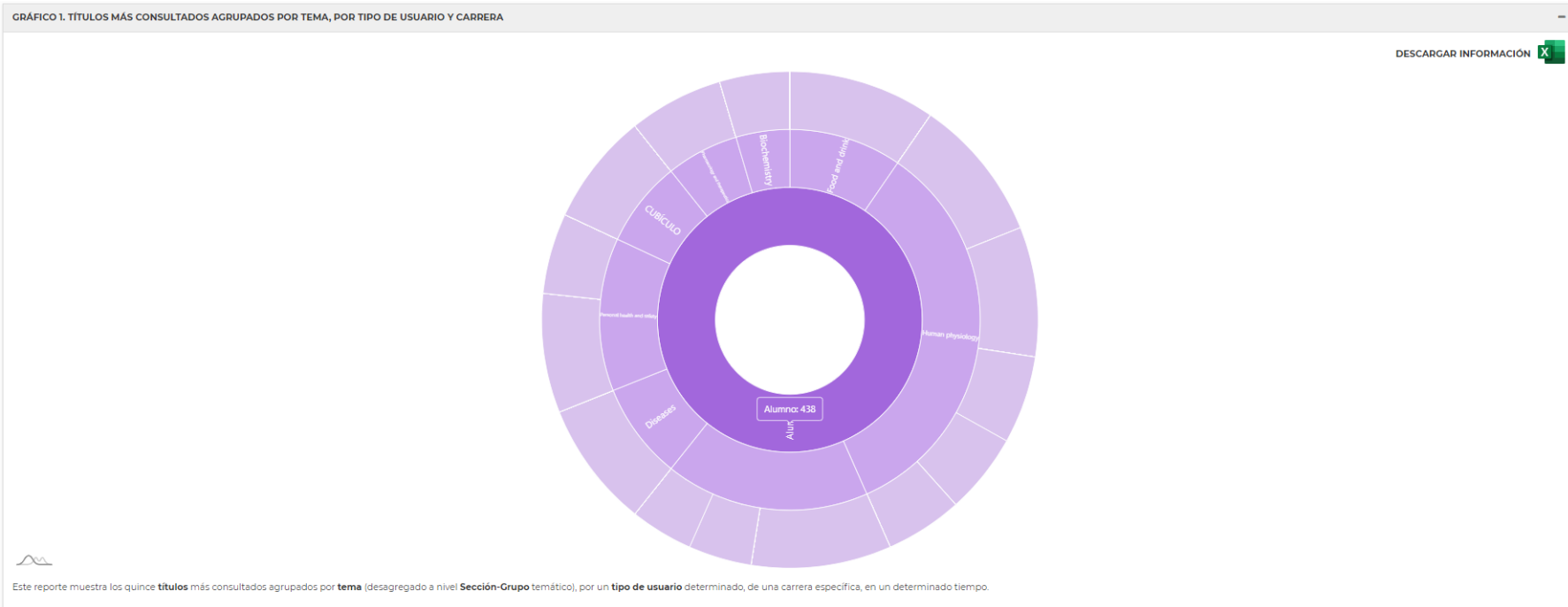
**SALIDAS DE LA CONSULTA:** Esta consulta tiene la posibilidad de descargarse en formato Excel para visualizar los datos de manera tabular, como se muestra en la Tabla 34.

**Tabla 34. Títulos más consultados por tipo de usuario y tema**

Tipo de usuario	Sección - Grupo	Título	Consultas
Alumno	Food and drink	Tablas de uso práctico de los alimentos de mayor consumo Miriam Muñoz edición Adolfo Chávez Vill	42
Alumno	Human physiology	Tratado de fisiología médica Arthur C. Guyton, John E. Hall; traducción Gea Consultoría Editoria	41
Alumno	Human anatomy, cytology, histology	Tratado de anatomía humana Fernando Quiroz Gutiérrez; colaboración Enrique Acosta [y otros 4]	40
Alumno	Human physiology	El ABCD de la evaluación del estado de nutrición Araceli Suverza Fernández, Karime Haua Navarro	37
Alumno	Diseases	Lo esencial en metabolismo y nutrición Amber Appleton, Olivia Vanbergen	36
Alumno	Personal health and safety	Manual de fórmulas y tablas para la intervención nutricional María Elena Palafox López, José	34
Alumno	CUBÍCULO	Cubículos de Estudio Grupales Sistema Universitario de Bibliotecas Académicas	32
Alumno	Pharmacology and therapeutics	Nutrición, diagnóstico y tratamiento Sylvia Escott Stump; traducción Martha Elena Araiza	27
Alumno	Human physiology	Nutrición clínica María Elena Téllez Villagómez	25
Alumno	Human physiology	Krause. Dietoterapia L. Kathleen Mahan, Sylvia Escott Stump, Janice L. Raymond	23
Alumno	Personal health and safety	Nutriología médica edición Esther Casanueva [y otros]	23
Alumno	Human physiology	Harper bioquímica ilustrada Robert K. Murray, Daryl K. Granner, Víctor W. Rodwell; traducción Víc	22
Alumno	Biochemistry	Bioquímica las bases moleculares de la vida Trudy McKee, James R. McKee; traducción Martha Elena A	20
Alumno	Human anatomy, cytology, histology	Moore. Anatomía con orientación clínica Keith L. Moore, Arthur F. Dalley, Anne M. R. Agur; traduc	18
Alumno	Human anatomy, cytology, histology	Histología texto y atlas. Correlación con biología celular y molecular Michael H. Ross, Wojciech	18

Fuente: Elaboración propia.

Para la visualización de esta consulta se utilizó un gráfico de tipo jerárquico, denominado gráfico de rayos solares detallados (*drill-down sunburst chart*). La Figura 53 muestra el gráfico desplegado por la interfaz de consulta.



**Figura 53. Títulos más consultados agrupados por tema, por tipo de usuario y carrera**

FUENTE: Elaboración propia.

**INTERPRETACIÓN:** Las salidas de la consulta 1 (gráfico y tabla en Excel) permitieron constar que los temas más consultados por los estudiantes de la Licenciatura en Nutrición durante el último semestre de 2019<sup>64</sup> fueron:

1. *Human physiology*: 148 préstamos
2. *Human anatomy, cytology, histology*: 76 préstamos
3. *Personal health and safety*: 57 préstamos
4. *Food and drink*: 42 préstamos
5. *Diseases*: 36 préstamos
6. CUBÍCULO: 32 préstamos
7. *Pharmacology and therapeutics*: 27 préstamos
8. *Biochemistry*: 20 préstamos

En total, de los 15 títulos con mayor circulación entre este tipo de usuario y carrera fueron realizados 438 préstamos. Esta consulta es útil para conocer cuáles son los títulos más solicitados según el tipo de usuario analizado, así como los temas de mayor interés para éste. Sirve como un apoyo para actividades de desarrollo de colecciones, pues permiten saber qué tipo de usuarios son los que hacen uso de la colección y cuáles son los principales temas que son requeridos por éstos.

## **CONSULTA 2. Transacciones realizadas por bibliotecario**

**NARRATIVA:** Este reporte muestra las transacciones (préstamos, renovaciones, devoluciones, multas cobradas, multas canceladas, errores de devolución) realizadas por un bibliotecario específico en un tiempo determinado.

**¿QUÉ SE CONSULTA?** TRANSACCIONES por BIBLIOTECARIO: número de PRÉSTAMOS, número de RENOVACIONES, número de DEVOLUCIONES, número de MULTAS COBRADAS, número de MULTAS CANCELADAS, número de ERRORES DE DEVOLUCIÓN DE MDOC, en una temporalidad determinada (TIEMPO: rango de fechas, años, semestres)

**¿CÓMO SE CONFIGURA ESTA CONSULTA?** La exploración de los datos comienza con la determinación de un espacio temporal específico (TIEMPO: años o semestres),

---

<sup>64</sup> Derivados de los 15 títulos con mayor circulación entre este tipo de usuario y esta carrera.

posteriormente se selecciona el BIBLIOTECARIO cuyas transacciones se explorarán. En el eje de las *y* se indicaría el porcentaje de las transacciones, en el eje de las *x* irían los años o semestres, cada barra correspondería a un año o semestre y estaría a su vez dividida en los porcentajes acumulados de las transacciones (PRÉSTAMOS, RENOVACIONES, DEVOLUCIONES, MULTAS COBRADAS, MULTAS CANCELADAS, ERRORES DE DEVOLUCIÓN DE MDOC) llevadas a cabo por el BIBLIOTECARIO bajo análisis.

**EJEMPLO DE CONSULTA:** ¿Cuáles fueron las transacciones realizadas por el bibliotecario DSARMANDO (número de préstamos, número de renovaciones, número de devoluciones, número de multas cobradas, número de multas canceladas y número de errores de devolución del MDoc) durante 2018 y 2019?

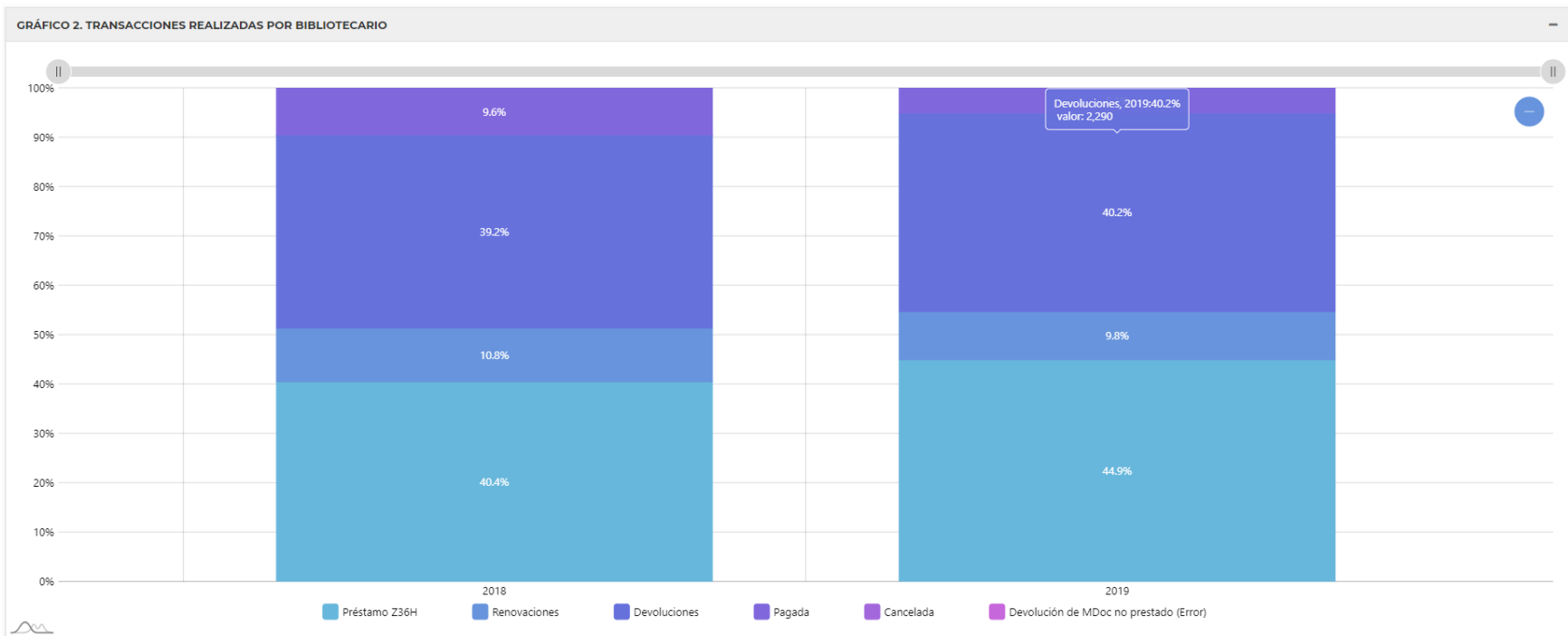
**SALIDAS DE LA CONSULTA:** Este reporte tiene formato de salida tabular en Excel, el cual se muestra en la Tabla 35.

**Tabla 35. Transacciones realizadas por bibliotecario**

<b>Año</b>	<b>Tipo de transacción</b>	<b>Cantidad</b>
2018	Préstamo Z36H	533
2018	Renovaciones	143
2018	Devoluciones	516
2018	Pagada	126
2019	Préstamo Z36H	2559
2019	Devoluciones	2290
2019	Renovaciones	557
2019	Pagada	293

Fuente: Elaboración propia.

Para esta consulta se empleó el gráfico interactivo de columnas 100% apiladas (*100% Stacked Column Chart*) que se muestra en la Figura 54.



**Figura 54. Transacciones realizadas por bibliotecario**

FUENTE: Elaboración propia.

**INTERPRETACIÓN:** Las salidas de esta consulta presentan las actividades realizadas por el bibliotecario seleccionado durante los años 2018 (1,318 transacciones) y 2019 (5,699 transacciones), en los cuales es posible ver un incremento de más del 400% con respecto a las actividades realizadas en el año previo. Proporcionalmente, el porcentaje de actividades realizadas (préstamos, renovaciones, devoluciones y atención del pago de multas) es muy similar entre años. Es notoria la ausencia de errores de devolución de parte del bibliotecario. Esta consulta sirve para realizar un monitoreo de las actividades realizadas por el personal durante sus periodos de actividad en la biblioteca, sobre todo para verificar si se presentan errores de devolución de MDoc que no fue prestado.

### **CONSULTA 3. Actualidad de la colección del MDoc por área temática**

**NARRATIVA:** Este reporte muestra la concentración de MDoc que existe en la BDESSalud de acuerdo con su año de publicación, ubicado en una escala temporal que abarca cada año del 2010 al 2019, por área temática.

**¿QUÉ SE CONSULTA?** Número de MDoc desagregado por TEMA (al nivel Sección/Grupo) y filtrado por TIEMPO (Año de publicación del MDoc) posicionado en una escala temporal del 2010 al 2019.

**¿CÓMO SE CONFIGURA ESTA CONSULTA?** La exploración de los datos comienza con la determinación de los TEMAS (al nivel Sección/Grupo) a revisar y el MDoc que pertenece a éstos. El MDoc se organiza (sumándose, en caso necesario) por año de publicación del MDoc. En el eje **x** se coloca la escala temporal anual del 2010 al 2019, en el eje **y** se señalan los TEMAS (al nivel Sección-Grupo) elegidos por el usuario. Los MDoc existentes en la BDESSalud se agrupan, de acuerdo con su CLASIFICACIÓN, en el nivel Sección/Grupo que les corresponda. Asimismo, se suman a aquellos que compartan su mismo año de publicación para determinar la columna que les corresponde (en el caso del material publicado antes del 2010, serán sumados a los valores de la columna 2010). A mayor número de MDoc en un año específico, mayor saturación de color en la celda correspondiente.



**EJEMPLO DE CONSULTA:** Dado que la BDESSalud atiende a usuarios especializados en el área de las Ciencias de la Salud, se delimitó una consulta del MDoc que posee perteneciente a los siguientes temas (a nivel Sección/Grupo) agrupado por año del MDoc, en el periodo 2010-2019:

1. *Medicine & health;*
2. *Human anatomy, cytology, histology;*
3. *Human physiology;*
4. *Personal health and safety;*
5. *Forensic medicine; incidence of injuries, wounds, disease, public preventive medicine*
6. *Pharmacology and therapeutics;*
7. *Diseases;*
8. *Surgery, regional medicine, dentistry, ophthalmology, otology, audiology;*
9. *Gynecology, obstetrics, pediatrics, geriatrics;*
10. *Sanitary engineering*

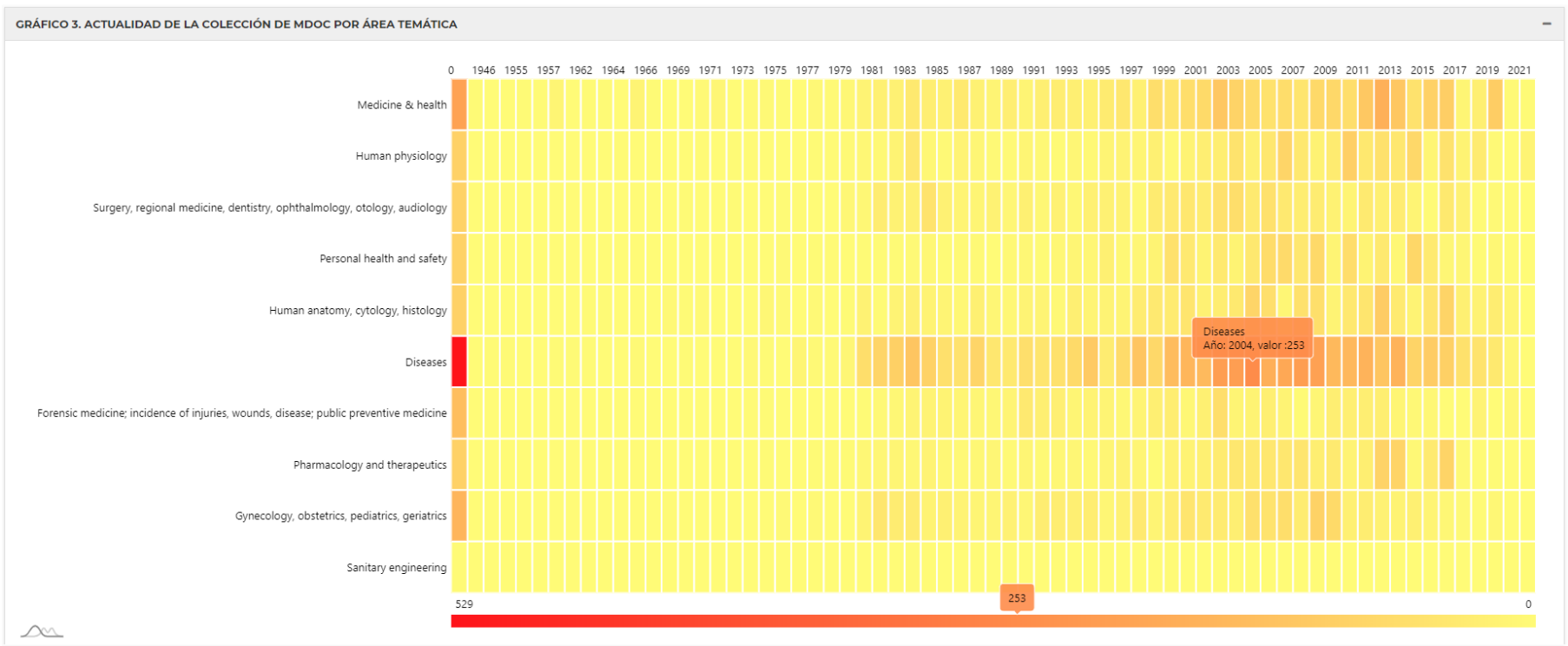
**SALIDAS DE LA CONSULTA:** Esta consulta tiene la posibilidad de obtener un formato tabular en Excel, cuyo extracto se muestra en la Tabla 36.

**Tabla 36. Actualidad de la colección del MDoc por área temática [fragmento]**

<b>Año MDoc</b>	<b>Sección/Grupo temático</b>	<b>Cantidad</b>
2021	Human physiology	14
2021	Medicine & health	2
2020	Forensic medicine; incidence of injuries, wounds, disease; public preventive medicine	8
2020	Medicine & health	4
2020	Surgery, regional medicine, dentistry, ophthalmology, otology, audiology	2
2020	Human anatomy, cytology, histology	12
2020	Diseases	10
2020	Personal health and safety	1
2019	Medicine & health	103
2019	Pharmacology and therapeutics	2
2019	Gynecology, obstetrics, pediatrics, geriatrics	7
2019	Human physiology	4
2019	Human anatomy, cytology, histology	30
2019	Surgery, regional medicine, dentistry, ophthalmology, otology, audiology	4
2019	Personal health and safety	8
2019	Diseases	41
2018	Diseases	54
2018	Human anatomy, cytology, histology	19
2018	Gynecology, obstetrics, pediatrics, geriatrics	1
2018	Medicine & health	27
2018	Personal health and safety	2
2018	Surgery, regional medicine, dentistry, ophthalmology, otology, audiology	2
2018	Pharmacology and therapeutics	1
2018	Human physiology	35
2018	Forensic medicine; incidence of injuries, wounds, disease; public preventive medicine	17

Fuente: Elaboración propia.

Para visualizar esta consulta se utilizó el gráfico de mapa de calor (*Heat Map with Legend*) que se muestra en la Figura 55.



**Figura 55. Actualidad de la colección de MDoc por área temática**

FUENTE: Elaboración propia.

A partir de la salida tabular de este reporte fue posible elaborar la Tabla 37, que muestra un resumen de la colección de la BDESSalud con respecto a los temas explorados.

**Tabla 37. Actualidad del MDoc de acuerdo con las áreas temáticas exploradas**

#	Área temática	MDoc actualizado (2016-2021)	MDOC DESACTUALIZADO (1946-2015)	MDoc sin información	TOTAL
1	Diseases	261	4,297	529	5,087
2	Medicine & health	251	2,031	201	2,483
3	Human anatomy, cytology, histology	174	1,124	95	1,393
4	Human physiology	126	954	95	1,175
5	Pharmacology and therapeutics	93	832	95	1,020
6	Surgery, regional medicine, dentistry, ophthalmology, otology, audiology	71	982	93	1,146
7	Personal health and safety	32	882	93	1,007
8	Forensic medicine; incidence of injuries, wounds, disease; public preventive medicine	25	652	137	814
9	Gynecology, obstetrics, pediatrics, geriatrics	20	1,148	158	1,326
10	Sanitary engineering	-	4	1	5
		1,053	12,906	1,497	15,456

Fuente: Elaboración propia

**INTERPRETACIÓN:** En la tabla de resumen obtenida de la salida tabular de la Consulta 3 es posible ver que *Diseases* es el área temática con mayor número de MDocs (261) con año de publicación reciente, seguida de *Medicine & health* (251) y *Human anatomy, cytology, histology* (174). Sin embargo, tanto *Diseases* (4,297) como *Medicine & health* (2,031) tienen el mayor número de MDoc desactualizados, seguidas de *Gynecology, obstetrics, pediatrics, geriatrics* (1,148) y *Human anatomy, cytology, histology* (1,124). Es importante notar que 1,497 MDocs no cuentan con la información referente al año de publicación, lo cual impidió su correcta agrupación tanto en el gráfico como en la tabla de salida.

Al observar la gráfica de calor es posible ver que la mayor cantidad de MDoc la concentra el tema de *Diseases*, seguido de *Medicine & health* y *Human anatomy, cytology, histology*. En cuanto a las adquisiciones más recientes de MDoc actualizado, éstas han sido en los temas de *Human physiology* y *Medicine & health*. Un área muy poco atendida ha sido *Sanitary engineering*, la cual tiene únicamente cinco ítems, de los cuales cuatro están desactualizados y uno no cuenta con información referente al año de publicación del MDoc.

Esta consulta es útil en las tareas de Desarrollo de Colecciones y sirve para conocer el grado de actualización del MDoc por área temática, tanto para conocer cuáles son las áreas debidamente actualizadas, como aquellas áreas débiles de la colección que necesitan reforzarse.

#### **CONSULTA 4. Tipo de adquisición del MDoc solicitado en préstamo por temporalidad**

**NARRATIVA:** Este reporte muestra el porcentaje acumulado del tipo de adquisición de los MDoc solicitados en préstamo por año.

**¿QUÉ SE CONSULTA?** TIPO DE ADQUISICIÓN del MDOC solicitado en PRÉSTAMO en la BDESSalud en el periodo de 2010-2019 (TIEMPO).

**¿CÓMO SE CONFIGURA ESTA CONSULTA?** La exploración de los datos comienza con la determinación de un espacio temporal específico (TIEMPO: 2010-2019). En el gráfico, el eje de las *x* presenta el rango temporal seleccionado y el eje de las *y* el porcentaje acumulado de PRÉSTAMOS del MDoc segmentado por el TIPO DE ADQUISICIÓN (con la opción de mostrar también el conteo de cada categoría).

**EJEMPLO DE CONSULTA:** ¿Cuál es el tipo de adquisición del MDoc solicitado en préstamo en la BDESSalud durante el periodo de 2010-2019?

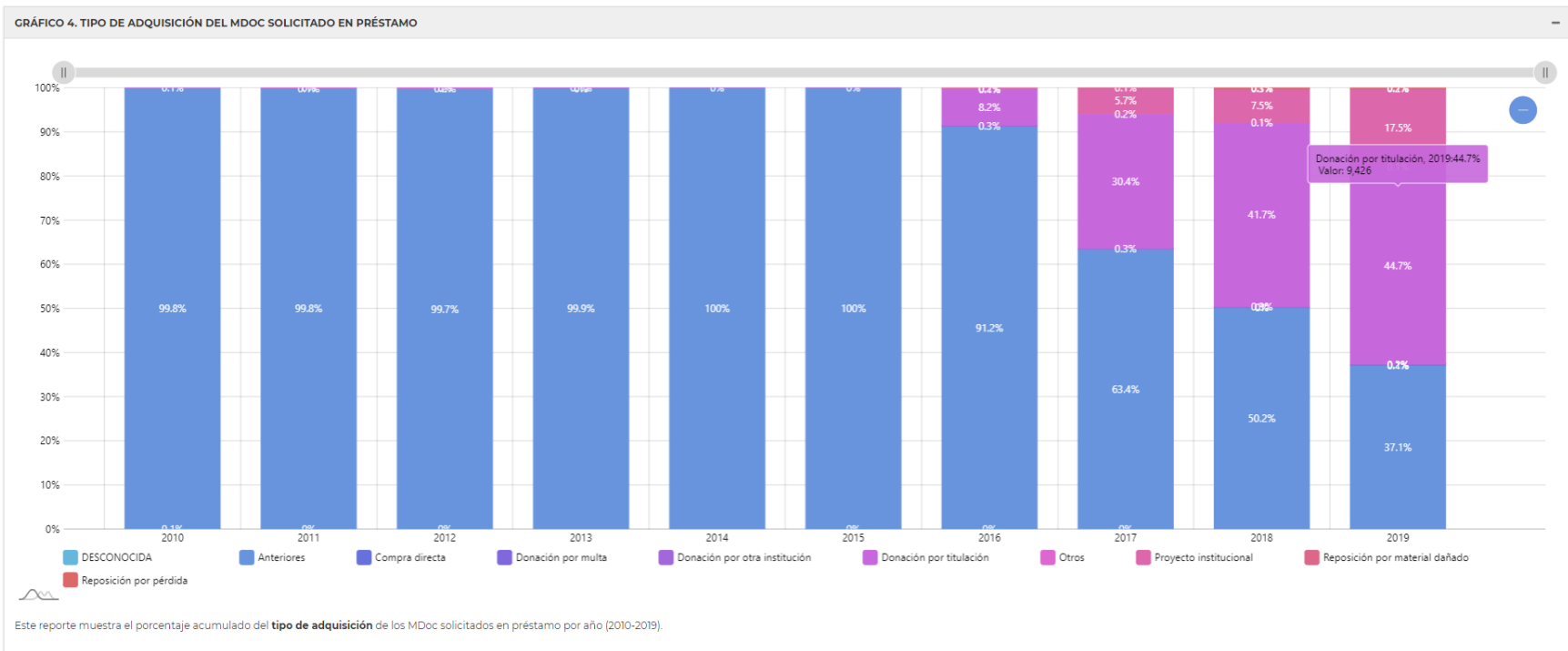
**SALIDAS DE LA CONSULTA:** Esta consulta tiene posibilidad de descarga del Excel para visualizar la información de manera tabular, como se muestra en la Tabla 38.

**Tabla 38. Tipo de adquisición del material solicitado en préstamo (2010-2019)**

<b>Año</b>	<b>Tipo de adquisición</b>	<b>Cantidad</b>
2010	DESCONOCIDA	19
2010	Anteriores	18,382
2010	Donación por titulación	14
2011	DESCONOCIDA	14
2011	Anteriores	38,877
2011	Donación por titulación	51
2011	Otros	16
2012	DESCONOCIDA	15
2012	Anteriores	36,675
2012	Donación por titulación	78
2012	Otros	14
2013	Anteriores	32,538
2013	Donación por otra institución	1
2013	Donación por titulación	21
2014	Anteriores	30,139
2014	Donación por titulación	6
2015	DESCONOCIDA	7
2015	Anteriores	28,237
2015	Donación por titulación	5
2016	DESCONOCIDA	9
2016	Anteriores	24,086
2016	Donación por otra institución	80
2016	Donación por titulación	2,157
2016	Otros	20
2016	Proyecto institucional	63
2017	DESCONOCIDA	8
2017	Anteriores	16,221
2017	Donación por otra institución	67
2017	Donación por titulación	7,775
2017	Otros	45
2017	Proyecto institucional	1,468
2017	Reposición por material dañado	13
2018	Anteriores	10,788
2018	Donación por multa	1
2018	Donación por otra institución	54
2018	Donación por titulación	8,958
2018	Otros	13
2018	Proyecto institucional	1,602
2018	Reposición por material dañado	64
2018	Reposición por pérdida	13
2019	Anteriores	7,830
2019	Compra directa	16
2019	Donación por otra institución	42
2019	Donación por titulación	9,426
2019	Otros	13
2019	Proyecto institucional	3,698
2019	Reposición por material dañado	32
2019	Reposición por pérdida	26

Fuente: Elaboración propia

Para esta consulta se empleó el gráfico interactivo de columnas 100% apiladas (100% *Stacked Column Chart*) mostrado en la Figuras 56.



**Figura 56. Tipo de adquisición del MDoc solicitado en préstamo**

FUENTE: Elaboración propia.



**INTERPRETACIÓN:** Este gráfico muestra cuál es el tipo de adquisición de los materiales solicitados en préstamo durante el periodo 2010-2019 en la BDESSalud. En esta consulta es evidente que a partir del 2016 los MDoc donados por titulación comienzan a desplazar a los materiales cuya vía de adquisición no se estipuló en el sistema. Lo anterior puede interpretarse en el sentido de que, al carecer de presupuesto asignado para adquisición de MDoc actual para la colección, se comenzó a priorizar la solicitud de donaciones a los egresados que se titulaban, los materiales más actualizados corresponden a tales donaciones y en la búsqueda de los usuarios por los materiales más nuevos, los préstamos aumentan. Esta consulta es útil para saber cuál es el tipo de adquisición que prevalece en los MDoc más solicitados por los usuarios.

#### **CONSULTA 5. Multas canceladas por bibliotecario, según el tipo de usuario y carrera**

**NARRATIVA:** Este reporte señala el número de multas canceladas por bibliotecarios en la BDESSalud durante un periodo específico, indicando el tipo de usuario al que le fue cancelada la multa y la carrera a que éste pertenece.

**¿QUÉ SE CONSULTA?** MULTAS CANCELADAS según el TIPO DE USUARIO y CARRERA durante un periodo específico (TIEMPO) por bibliotecarios de la BDESSalud.

**¿CÓMO SE CONFIGURA ESTA CONSULTA?** Comenzamos el análisis de datos seleccionando en el tablero de consulta en el filtro de ESTADO DE MULTA, la opción MULTA CANCELADA, indicamos la opción temporal del último semestre de 2019 (TIEMPO). En la gráfica elegida:

- La burbuja principal constituye el total de multas canceladas en el periodo especificado, al colocar el cursor sobre la burbuja muestra el número de multas cancelada en esta categoría.
- La burbuja secundaria corresponde al bibliotecario que llevó a cabo la cancelación de la multa, al colocar el cursor sobre la burbuja muestra el número de multas cancelada en esta categoría.

- La burbuja terciaria indica el tipo de usuario al que le fue cancelada la multa, así como la carrera a que éste pertenece. Al colocar el cursor sobre la burbuja muestra el número de multas cancelada en esta categoría.

**EJEMPLO DE CONSULTA:** ¿Cuál es el número de multas canceladas por bibliotecarios de la BDESSalud durante el último semestre de 2019? ¿A qué tipo de usuario le fueron canceladas tales multas? ¿A qué carreras pertenecen estos usuarios?

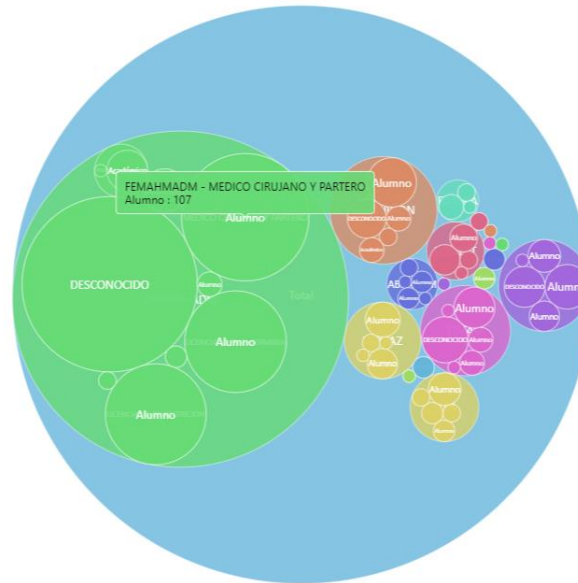
**SALIDAS DE LA CONSULTA:** Esta consulta permite la descarga del formato Excel para visualizar los datos de manera tabular, como se muestra en la Tabla 39.

**Tabla 39. Multas canceladas por bibliotecario, según el tipo de usuario y carrera**

Bibliotecario	Carrera	Tipo de usuario	Multas canceladas
FEMAHMADM	DESCONOCIDO	DESCONOCIDO	203
FEMAHMADM	MEDICO CIRUJANO Y PARTERO	Alumno	107
FEMAHMADM	LICENCIADO EN ENFERMERIA	Alumno	68
FEMAHMADM	LICENCIADO EN NUTRICION	Alumno	67
DSJQUIÑON	MEDICO CIRUJANO Y PARTERO	Alumno	16
DSFABI	DESCONOCIDO	DESCONOCIDO	14
FMFRESCAS	LICENCIADO EN ENFERMERIA	Alumno	13
FEMAHMADM	INGENIERO BIOMEDICO	Alumno	12
DSFABI	MEDICO CIRUJANO Y PARTERO	Alumno	11
FEMAHMADM	PERSONAL ACADEMICO	Académico	11
FMFRESCAS	DESCONOCIDO	DESCONOCIDO	11
DSJQUIÑON	DESCONOCIDO	DESCONOCIDO	10
DSKPAZ	MEDICO CIRUJANO Y PARTERO	Alumno	8
FMFRESCAS	MEDICO CIRUJANO Y PARTERO	Alumno	7
LUIS	MEDICO CIRUJANO Y PARTERO	Alumno	7
DSIVETT	DESCONOCIDO	DESCONOCIDO	6
DSKPAZ	LICENCIADO EN ENFERMERIA	Alumno	6
FMFRESCAS	LICENCIADO EN NUTRICION	Alumno	6
DSIVETT	MEDICO CIRUJANO Y PARTERO	Alumno	5
DSFABI	ENFERMERIA GENERAL CON BACHILLERATO	Alumno	4
DSFABI	LICENCIADO EN ENFERMERIA	Alumno	4
DSJQUIÑON	LICENCIADO EN ENFERMERIA	Alumno	4
DSJQUIÑON	PERSONAL ACADEMICO	Académico	4
FEMAHMADM	ENFERMERIA GENERAL CON BACHILLERATO	Alumno	4
FINORA	DESCONOCIDO	DESCONOCIDO	4

FUENTE: Elaboración propia.

Para el despliegue visual de esta consulta se empleó el gráfico jerárquico de círculos empaquetados (*Packed Circle Chart*) que se muestra en la Figura 57.



Este reporte señala el número y porcentaje de **multas canceladas** por **bibliotecarios** en la DESSalud durante un periodo específico, indicando el **tipo de usuario** al que le fue cancelada la multa y la **carrera** a que éste pertenece.

**Figura 57. Multas canceladas por bibliotecario, según el tipo de usuario y carrera**

FUENTE: Elaboración propia.

**INTERPRETACIÓN:** Esta consulta permite monitorear cuáles fueron los bibliotecarios que han realizado cancelaciones de multas, así como saber a qué tipo de usuario se le dio ese beneficio de manera más frecuente y a qué carrera pertenece. Es un complemento para los reportes de multas existentes en el LogAleph de la BDESSalud.

#### **CONSULTA 6. Tipo de devolución vencida por tipo de usuario, según el tema del MDoc**

**NARRATIVA:** Este reporte muestra el tipo de devolución vencida por tipo de usuario que efectuó la devolución, dividido posteriormente por el tema (a nivel Sección/Grupo) a que pertenece el MDoc devuelto, en un tiempo determinado.

**¿QUÉ SE CONSULTA?** TIPO DE DEVOLUCIÓN vencida por TEMA (a nivel Sección/Grupo), por TIPO DE USUARIO y TEMPORALIDAD

**¿CÓMO SE CONFIGURA ESTA CONSULTA?** Se delimita desde el tablero de consulta la opción “Vencidas” del filtro TIPO DE DEVOLUCIÓN, así como la TEMPORALIDAD (2010-2019) en la que fueron realizados estas transacciones.

Gráfica a nivel 1: Se muestra el total de TIPO DE DEVOLUCION “vencida” del MDOC por TIPO-USUARIO que efectuó la DEVOLUCION (al pasar el cursor indica el número de usuarios y/o porcentaje).

Gráfica a nivel 2. Se muestra el TEMA (a nivel Sección/Grupo) y al pasar el cursor indica el número de devoluciones vencidas y/o porcentaje por tema.

**EJEMPLO DE CONSULTA:** ¿Cuál es el número de devoluciones vencidas durante el periodo 2010-2019 en la BDESSalud, qué tipo de usuarios incurrieron en tal transacción y cuál fue el tema de los materiales devueltos vencidos?

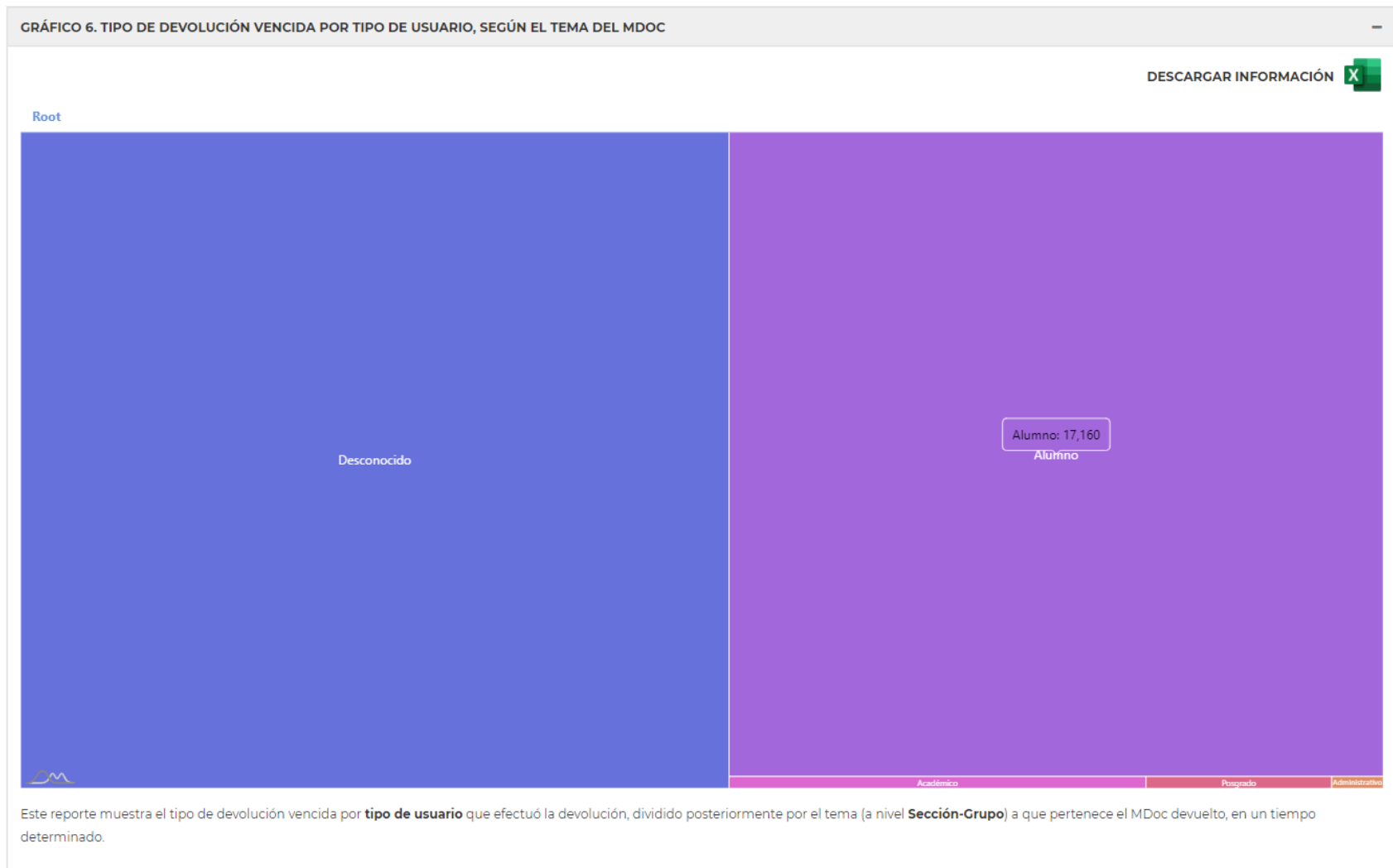
**SALIDAS DE LA CONSULTA:** Esta consulta contó con una salida tabular mostrada en la Tabla 40.

**Tabla 40. Devoluciones vencidas por tipo de usuario, según el tema del MDoc**

Tipo de usuario	Sección / Grupo temático	Cantidad de devoluciones vencidas
Desconocido	Diseases	6,255
Alumno	Diseases	3,759
Alumno	Human anatomy, cytology, histology	3,102
Alumno	Human physiology	2,876
Desconocido	Human anatomy, cytology, histology	2,845
Desconocido	Human physiology	2,479
Alumno	Pharmacology and therapeutics	2,001
Alumno	Medicine & health	1,426
Desconocido	Pharmacology and therapeutics	1,223
Desconocido	Gynecology, obstetrics, pediatrics, geriatrics	1,056
Desconocido	Medicine & health	988
Alumno	Biochemistry	846
Desconocido	Surgery, regional medicine, dentistry, ophthalmology, otology, audiology	828
Desconocido	Biochemistry	728
Alumno	Personal health and safety	723
Desconocido	Personal health and safety	671
Alumno	Gynecology, obstetrics, pediatrics, geriatrics	481
Alumno	Surgery, regional medicine, dentistry, ophthalmology, otology, audiology	450
Alumno	Food and drink	265
Desconocido	Genetics and evolution	228
Desconocido	Forensic medicine; incidence of injuries, wounds, disease; public preventive medicine	193
Desconocido	Food and drink	185
Desconocido	Food technology	174
Alumno	Physiology and related subjects	135
Alumno	Knowledge	130
Desconocido	Physiology and related subjects	121
Desconocido	Knowledge	118
Alumno	Genetics and evolution	115
Desconocido	General management	112
Alumno	Chemistry and allied sciences	97
Alumno	Food technology	97
Alumno	Forensic medicine; incidence of injuries, wounds, disease; public preventive medicine	93
Desconocido	Chemistry and allied sciences	81
Alumno	Occupational ethics	81
Desconocido	Biology	62

Fuente: Elaboración propia.

Para la visualización de esta consulta se empleó el gráfico jerárquico de mapa de árbol detallado (*Drill-down tree map*) que se muestra en las Figuras 58 y 59.



**Figura 58. Devoluciones vencidas por tipo de usuario, según tema del MDoc, 1ª capa**

FUENTE: Elaboración propia.



**Figura 59. Devoluciones vencidas por tipo de usuario, según tema del MDoc, 2ª capa**

FUENTE: Elaboración propia.

**INTERPRETACIÓN:** En esta consulta es notoria la presencia de usuarios que carecen de información inherente (en este caso faltan los datos asociados a su tipo de usuario) quienes concentran 18,930 devoluciones vencidas (52% del total). Posteriormente, el tipo de usuario que concentra la mayor parte de las devoluciones y el cual es factible de ser identificado es la categoría de Alumno, con 17,160 devoluciones vencidas (47.13%). Le siguen Académicos con 202 devoluciones vencidas (0.55%), Posgrado con 90 devoluciones vencidas (0.25%) y Administrativos con 25 devoluciones vencidas (0.07%). De los estudiantes, los temas con mayor número de devoluciones vencidas son *Diseases*: (3,759), *Human anatomy, cytology, histology* (3,102) y *Human physiology* (2,876), *Pharmacology and therapeutics* (2,001) y *Medicine & health* (1,426).

Esta consulta resulta de utilidad para saber cuáles son los temas más susceptibles de devolverse de manera tardía, así como cuáles son los tipos de usuarios más proclives en caer en dicha práctica.

### **3.2 Resultados del ejercicio de bibliominería**

Para clasificar a los usuarios considerando los atributos, criterios o características para el estudio se seleccionó la técnica de agrupamiento *K-Means*. En este caso, se consideraron las características inherentes a los usuarios, así como la información transaccional que éstos generaron al interactuar con el sistema.

Los hallazgos muestran que, luego de la aplicación del algoritmo de agrupación, fue posible crear cuatro perfiles de usuarios que poseen características distintivas de acuerdo con ciertos patrones conformados por sus características inherentes y los comportamientos informativos detectados en el uso del MDoc de la BDESSalud.

Los principales resultados se pueden resumir en la Tabla 41, la cual muestra las principales características resultantes de la segmentación en cuatro grupos.



**Tabla 41. Principales características de los cuatro grupos identificados**

<b>Grupos</b>	<b>Porcentaje de usuarios</b>	<b>Carrera</b>	<b>Facultad</b>	<b>Porcentaje de mujeres en el grupo</b>	<b>Promedio mensual de préstamos</b>	<b>Promedio de préstamos en meses con actividad</b>	<b>Antigüedad promedio del MDoc consultado</b>	<b>Promedio de días de retraso por préstamo</b>	<b>Promedio de % de devoluciones tardías</b>	<b>Promedio de temas generales distintos consultados</b>
<b>Grupo 1</b>	84%	Médico Cirujano Partero	Facultad de Medicina y Ciencias Biomédicas	100%	2.0	2.7	12.7	0.8	0.1	4.4
<b>Grupo 2</b>	81%	Licenciado en Enfermería	Facultad de Enfermería y Nutriología	78%	1.4	2.4	16.1	1.2	0.1	4.1
<b>Grupo 3</b>	83%	Médico Cirujano Partero	Facultad de Medicina y Ciencias Biomédicas	0%	1.9	2.7	13.1	3.3	0.2	4.4
<b>Grupo 4</b>	94%	Licenciado en Nutrición	Facultad de Enfermería y Nutriología	80%	1.4	2.5	14.8	0.5	0.1	5.4

Fuente: Elaboración propia.

Se destaca que los grupos 1 y 3 se incluyen a personas en su mayoría de la carrera de Médico Cirujano Partero, de la Facultad de Medicina y Ciencias Biomédicas; en donde el grupo 1 se distingue en que sólo lo constituyen mujeres usuarias, mientras que el grupo 3 está constituido sólo por hombres usuarios.

El grupo 2 lo constituyen principalmente personas de la carrera de Licenciado en Enfermería, y el grupo 4 lo constituyen en su mayoría personas de la carrera de Licenciado en Nutrición, ambos de la Facultad de Enfermería y Nutriología, y ambos grupos constituidos en su mayoría por mujeres<sup>65</sup>.

En el grupo 1 y 3 están los usuarios más intensivos en préstamos, con un promedio mensual de 2 y 1.9 préstamos por mes. También consultan el MDoc con una menor antigüedad promedio. Los académicos y usuarios con posgrado se encuentran en su mayoría en el grupo 1.

Por otra parte, el grupo 2 es el que tiene un menor promedio de préstamos en meses con actividad y consulta material con antigüedad promedio mayor, también tiene el menor número promedio de temas generales distintos consultados.

El grupo 3 consulta MDoc de menor antigüedad, aunque también tiene el mayor promedio de días de retraso por préstamo y el mayor porcentaje de devoluciones tardías<sup>66</sup>.

El grupo 4 tiene el menor promedio de retrasos por préstamo y son los que consultan más temas generales distintos.

En lo referente a las temáticas consultadas por los distintos grupos, todos ellos hacen uso intensivo de los temas "*Human physiology*" en primer lugar, seguido por "*Human anatomy, cytology, histology*". En particular el grupo 4 también hace uso intensivo del tema "*Personal health and safety*".

Al analizar la especialización en temas por parte de los distintos grupos (véase Tabla 42), se observa que el grupo 1 y 3 consultan temas similares, aunque el grupo 3 también

---

<sup>65</sup> Cabe señalar que el 67% de los usuarios son mujeres.

<sup>66</sup> Este dato puede indicar la necesidad de establecer comunicación con este grupo usuario en particular para buscar un mejor comportamiento en el uso del MDoc y reducir el número de devoluciones tardías.

se enfoca en "Gynecology, obstetrics, pediatrics, geriatrics" y "Pharmacology and therapeutics" y en menor medida "Surgery, regional medicine, dentistry, ophthalmology, otology, audiology".

**Tabla 42. Diferencias en temas consultados según el grupo de usuarios**

Grupos	Diferencias en temas más consultados						
<b>Grupo 1</b>	<i>Human anatomy, cytology, histology</i>	<i>Diseases</i>	<i>Biochemistry</i>	<i>Knowledge</i>	<i>Physiology and related subjects</i>		
<b>Grupo 2</b>	<i>Medicine &amp; health</i>	<i>Occupational ethics</i>					
<b>Grupo 3</b>	<i>Human anatomy, cytology, histology</i>	<i>Diseases</i>	<i>Biochemistry</i>	<i>Gynecology, obstetrics, pediatrics, geriatrics</i>	<i>Pharmacology and therapeutics</i>	<i>Physiology and related subjects</i>	<i>Surgery, regional medicine, dentistry, ophthalmology, otology, audiology</i>
<b>Grupo 4</b>	<i>Human physiology</i>	<i>Personal health and safety</i>	<i>Food and drink</i>	<i>Food technology</i>	<i>Chemistry and allied sciences</i>	<i>Forensic medicine; incidence of injuries, wounds, disease; public preventive medicine</i>	

Fuente: Elaboración propia.

El grupo 2 se destaca por consultar temas de "Medicine & health" y "Occupational Ethics".

El grupo 4 se destaca por consultar los temas de "Human physiology", "Personal health and safety", "Food and drink", "Chemistry and allied sciences" y "Forensic medicine; incidence of injuries, wounds, disease; public preventive medicine".

Los resultados obtenidos en la aplicación de la bibliominería en el DW de la BDESSalud pueden emplearse como apoyo para la delimitación de políticas de préstamo apropiadas, además de ser de utilidad para mejorar la comunicación con la comunidad usuaria y, de acuerdo con los perfiles identificados, mantener a los usuarios informados con respecto a la llegada de materiales actualizados o novedosos que pertenezcan a sus áreas temáticas de interés.

### 3.3 Modelo de bibliominería propuesto

Considerando la metodología delimitada en el Capítulo 2, su aplicación y los resultados obtenidos, es posible elaborar un esbozo gráfico en la Figura 60 del modelo de bibliominería integrado con estos elementos.

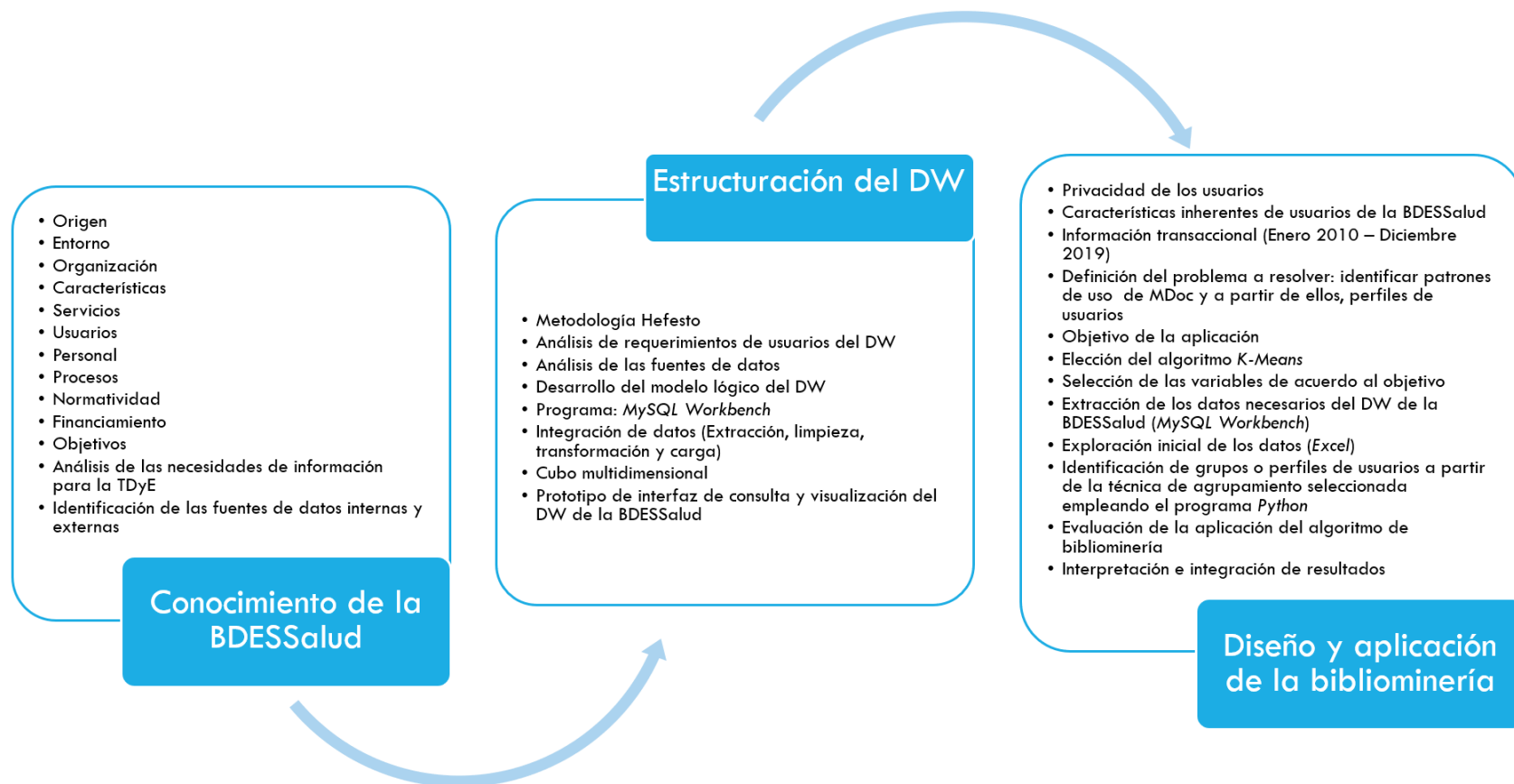


Figura 60. Esquema del modelo de bibliominería

FUENTE: Elaboración propia.

# CAPÍTULO 4. Discusión

En este apartado se realiza un ejercicio de contraste entre el planteamiento inicial y los objetivos del estudio, los resultados obtenidos en el desarrollo de la tesis y la teoría revisada durante su elaboración. En las distintas secciones que lo componen se presentan los argumentos en donde se visibiliza la postura de su autora ante los resultados obtenidos, exponiéndose y valorándose tanto los aportes como las limitantes del estudio.

## 4.1 Sobre los planteamientos iniciales de la tesis

En la parte introductoria de este trabajo se presentó el problema de investigación configurado a partir de la siguiente pregunta: **¿Cuáles son las estructuras y conjuntos de datos que se requieren para la identificación de patrones de uso que generen conocimiento significativo y en qué medida éstos contribuyen a los procesos de evaluación y toma de decisiones de las actividades bibliotecarias?**

Usualmente, las bibliotecas cuentan con sistemas especializados (SIGB's) que les permiten llevar a cabo sus actividades fundamentales y ofrecer los servicios de información requeridos por su comunidad usuaria. Sin embargo, estos sistemas suelen ofrecer opciones restringidas para efectuar análisis avanzados de los datos generados por las transacciones diarias y los datos acumulados derivados de tales actividades. El intentar llevar a cabo este tipo de tareas enfrenta al profesional de la información con el reto de lidiar con sistemas de informes obsoletos (Ball, 2019, p. 6), y/o predeterminados que no le permiten comprender a cabalidad el conocimiento potencial embebido en esos datos.

De acuerdo con Siguenza-Guzman, los sistemas tradicionales de gestión de bibliotecas son adecuados para satisfacer las necesidades de automatización de las bibliotecas en términos de actividades operativas y servicios físicos tradicionales. Sin embargo, [no lo

son para] promover la innovación y la creación de conocimiento para la toma de decisiones estratégicas, [pues] los sistemas tradicionales brindan funciones muy limitadas (Siguenza-Guzman, 2015, p. 140).

La BDESSalud opera empleando el SIGB Aleph, el cual constituye un producto heredado de ExLibris que alcanzó su punto máximo en ventas en 2015 y ha disminuido a medida que las bibliotecas universitarias y académicas se decantan por la implementación de ALMA y otros productos similares. “En los últimos años, las bibliotecas universitarias se han alejado radicalmente de los sistemas bibliotecarios integrados (ILS) orientados a las colecciones impresas, adoptando en su lugar plataformas de servicios bibliotecarios (LSP) diseñadas para gestionar todos los posibles formatos que integran una colección. Las instalaciones de productos ILS como Aleph [...], orientados a bibliotecas universitarias, están perdiendo interés, y los proveedores que ofrecen productos ILS adoptados por múltiples tipos de bibliotecas están viendo cómo estos sistemas se sustituyen por los nuevos productos en colegios y universidades” (Breeding, 2021). Sin embargo, dada la falta de presupuesto, es difícil que la BDESSalud opte por la adquisición de un ILS en un futuro próximo. En cuanto a la generación de reportes, al BDESSalud emplea el sistema LogAleph, que desafortunadamente ha presentado fallos constantes en los últimos años. Esto hace necesaria la búsqueda de opciones que le permitan optimizar el uso de los sistemas existentes y contar con insumos útiles para la evaluación y toma de decisiones.

El principal objetivo de esta tesis fue determinar las estructuras de datos, así como las metodologías y herramientas que propicien la identificación de patrones de uso que generen conocimiento significativo con el propósito de establecer su relación con los procesos de evaluación y la toma de decisiones en actividades bibliotecarias. Para ello se recurrió al estudio de las opciones ofrecidas por la Ciencia de datos para optimizar el análisis de los datos acumulados por la biblioteca, enfocándonos en el KDD y en particular, los DW y la bibliominería. La revisión de la literatura relativa al tema permitió conocer que una adecuada implementación de un proceso de bibliominería se ve beneficiada con el diseño y habilitación previa de un almacén de datos o DW, por lo cual se eligió la metodología Hefesto (Bernabeau y García, 2017) para estructurar el DW de

la BDESSalud. Autores como Gutiérrez y Barranco (2008, p. 135) y Raina y Krishnamurthy (2022, 173) indican que, aun cuando se pudieran procesar directamente los datos con algoritmos de MD y/o aprendizaje automático, el que los datos sean sometidos a una preparación previa podría resultar en modelos más efectivos y de mayor calidad. Así, la información que brinda un DW, de tipo retrospectivo, con la posibilidad de realizar análisis dinámicos en varios niveles de detalle, permite tener los datos (obtenidos de fuentes heterogéneas) limpios, transformados, integrados y organizados de manera tal que se facilita su minado posterior y la obtención de patrones más certeros.

En el caso de estudio tratado en esta tesis resultó realmente útil el contar con el DW, pues los datos contenidos en las fuentes, sobre todo de las tablas Oracle de Aleph, requirieron de un gran trabajo de depuración para obtener los insumos para alimentar el propio DW y, por ende, la creación de una vista minable consolidada, al permitir que los datos fueran, a partir de su limpieza e integración, más consistentes y completos. Lo anterior permite confirmar lo que señalan Raina y Krishnamurthy (2022, p. 173) cuando indican que “preparar los datos de varias maneras podría dar como resultado modelos más efectivos”. Esto fue patente, por ejemplo, al momento de armonizar los datos referidos al año de publicación y sumarlos a los datos referidos al tema del MDoc de la BDESSalud en el DW estructurado. Durante la revisión de las tablas proporcionadas por la Unidad de Sistemas fue posible detectar la falta de uso en el SUBA del campo del Aleph destinado a indicar el año del MDoc. La manera en que se obtiene este dato depende de las consultas SQL establecidas en la Unidad de Sistemas y desafortunadamente éstas no siempre logran obtener los datos precisos. Esto conlleva muchas limitantes a la hora de desarrollar estudios referentes al año del MDoc. Durante la estructuración del DW se hizo un importante esfuerzo por recuperar y depurar la mayor cantidad de datos posibles en este tema, gracias a lo cual fue posible la confección de la Consulta 3, Actualidad de la colección del MDoc por área temática (Tablas 36, 37 y Figura 55). Las salidas de esta consulta resultan particularmente útiles en las tareas de Desarrollo de Colecciones y, de no ser gracias al DW de la BDESSalud, no habría sido posible tener una consulta de este tipo, pues los temas del MDoc no están contemplados en ninguno de los reportes existentes al momento en el LogAleph.

El uso de la metodología Hefesto implicó el determinar cuáles eran las necesidades de información de los tomadores de decisiones de la biblioteca con el fin de seleccionar los datos para integrar el DW de manera adecuada. De esta manera se cumplió con lo establecido en el objetivo específico de “identificar las relaciones existentes entre los datos estadísticos de series cronológicas derivados de los procesos y servicios bibliotecarios y los requerimientos para la evaluación y toma de decisiones de una biblioteca universitaria, con el propósito de seleccionar los datos que integrarán el repositorio de datos diseñado con estos fines”. Este paso fue de suma importancia porque permitió el involucrar a actores de interés en la BDESSalud en la delimitación de objetivos para la evaluación que pudieran ser abordados a partir del proceso de bibliominería, lo cual también contribuyó a que estuvieran más abiertos a compartir información y sobre todo compartir sus experiencias en el trabajo que se lleva a cabo en la BDESSalud, aspectos de real interés para el desarrollo del proyecto. En este aspecto cabe destacar que hubo un momento durante la realización de esta tesis en que se estableció contacto con el principal teórico de la bibliominería: el Dr. Scott Nicholson, para indagar las posibles causas que motivaron su cambio radical de temática de la bibliominería al diseño de juegos educativos. El autor señaló lo siguiente: “Me detuve porque tenía una solución sin un problema. Sin una razón bien definida para la bibliominería, me resultó difícil lograr que los bibliotecarios quisieran dedicar tiempo y correr el riesgo de recopilar y limpiar datos”. Esto permitió afirmar que el tomar el tiempo necesario para definir las necesidades y problemáticas a que se enfrentan los tomadores de decisiones es un paso fundamental para el adecuado desarrollo de un proyecto de DW y bibliominería.

Los objetivos específicos referentes a “diseñar y estructurar un repositorio de datos orientado a la identificación de patrones de uso que brinden conocimiento significativo sobre procesos, tareas y operaciones unitarias de los servicios bibliotecarios”, así como “proponer un modelo de bibliominería con el fin de explicar las relaciones teóricas y metodológicas existentes entre la estructura de datos mensurables que se generan o pudieran generar en actividades bibliotecarias y su capacidad de aportar información útil en los procesos de evaluación y toma de decisiones de tales actividades” también fueron cumplidos de manera satisfactoria en el desarrollo de esta tesis, como queda



ampliamente documentado en el segundo y tercer capítulos de este escrito. En este punto, fue particularmente importante el resultado obtenido en el ejercicio de bibliominería en el cual fue posible delimitar los perfiles de cuatro tipos de usuarios asiduos de la BDESSalud (Véase Tabla 41. Principales características de los cuatro grupos identificados). Estos resultados permiten tener una mayor claridad con respecto a las características inherentes de los usuarios (en cuanto a su información demográfica disponible: sexo, tipo de usuario, carrera, facultad) y a su comportamiento informativo (en cuanto a sus patrones de uso de MDoc) de acuerdo con la información transaccional analizada (promedio mensual de préstamos, promedio de préstamos en meses con actividad, antigüedad del material consultado, promedio de días de retraso por préstamo, promedio de devoluciones tardías, promedio de temas generales distintos consultados, temas de interés), lo cual puede emplearse como apoyo para el desarrollo de colecciones, la delimitación de planes de adquisiciones de MDoc, el diseño de políticas de préstamo apropiadas, además de ser de utilidad para mejorar la comunicación con la comunidad usuaria y, de acuerdo con los perfiles identificados, mantener a los usuarios informados con respecto a la llegada de materiales actualizados o novedosos que pertenezcan a las áreas temáticas de su interés. Asimismo, “al proporcionar información sobre el rendimiento y la utilidad de la biblioteca como una unidad, la bibliominería puede brindar la justificación necesaria para obtener apoyo financiero e institucional de manera continua para sus operaciones” (Pal, 2011, p. 12-13). En este sentido, uno de los resultados obtenidos en la Consulta 3 del DW de la BDESSalud (referente a la actualidad de la colección de MDoc por área temática) es que gran parte de los materiales analizados están desactualizados, pues presentan una antigüedad mayor a los cinco años. Las consultas hechas al DW, así como el ejercicio de bibliominería, permitieron identificar que los usuarios tienden a consultar este tipo de material habitualmente. Este resultado puede constituir un apoyo para la BDESSalud a la hora de negociar la asignación de presupuesto que sea destinado para el adecuado desarrollo de colecciones. Recordando la recomendación de Litton (1974, p.35-36), es necesario que el SUBA cuente con un reglamento autorizado por el Consejo Universitario de la UACH en el cual se garantice el apoyo al sistema bibliotecario con respecto a una justa asignación presupuestal, lo cual facilitaría el adecuado desarrollo de sus servicios.

Los resultados del ejercicio de bibliominería se complementan bien con las consultas que han sido delimitadas hasta el momento para el aprovechamiento del DW de la BDESSalud, pues, a partir de los perfiles delimitados, es factible el enriquecer el conocimiento que se tiene de estos cuatro tipos de usuarios empleando las consultas disponibles. Sin embargo, una recomendación a futuro es el robustecer el DW a partir de la confección de nuevas consultas que atiendan a las necesidades de información que los tomadores de decisiones en la BDESSalud vayan presentando. Asimismo, pueden replantearse nuevas aplicaciones de bibliominería considerando las necesidades de los tomadores de decisiones, pues a decir de Nicholson, “[e]l proceso es de naturaleza cíclica: a medida que se descubren patrones, surgen más preguntas que iniciarán el proceso nuevamente. A medida que se exploren áreas adicionales de la biblioteca, el DW se volverá más completo, lo que facilitará mucho la exploración de otros problemas” (Nicholson, 2003a, p. 146).

El objetivo final de esta tesis fue el identificar qué datos no se generan y/o almacenan en el sistema, que podrían ser útiles para llevar a cabo otros estudios y obtener mayor conocimiento<sup>67</sup>. En este aspecto es posible declarar que las bases de datos que contienen información tanto de usuarios estudiantes, docentes, investigadores, administrativos o bibliotecarios<sup>68</sup> son de vital importancia para establecer categorías o delimitaciones sobre el tipo de usuarios, sus características demográficas, académicas y profesionales, entre otras. Si se desea efectuar una evaluación del impacto de las bibliotecas académicas y el cómo éste se relaciona a la calidad educativa en instituciones de educación superior, aparte de contar con la información relativa al funcionamiento de los centros de información, es necesario contar con información relacionada al desempeño académico de los estudiantes<sup>69</sup>, como pueden ser:

---

<sup>67</sup> Véase Anexo 6. Variables de interés vs variables disponibles en la BDESSalud

<sup>68</sup> Datos a los que no fue posible tener acceso para el desarrollo de esta investigación, como es patente en el Anexo 6.

<sup>69</sup> Por ejemplo, Villalobos y Portas Lagar realizaron en el 2013 un “análisis del ingreso a los recursos de información y su relación con los promedios académico en el ITESO”, para lo cual emplearon lo que denominaron una herramienta de descubrimiento de datos (o como lo denominan los autores, *data discovery*) que les permitió realizar un sinnúmero de combinaciones de datos de manera instantánea, a partir de la asociación de los datos disponibles a través de filtros por el campo deseado. Su principal hallazgo es que “sí hay una correlación entre estudiantes que ingresaron a los recursos de información y el promedio académico obtenido. [Empero,] es

- Puntuaciones de admisión de exámenes de admisión a posgrado.
- Puntuaciones en certificaciones.
- Puntuaciones obtenidas por profesores.
- Exámenes para evaluar pensamiento crítico, juicios, razonamiento analítico, solución de problemas, comunicación (habilidades de escritura) y toma de decisiones.
- Participación estudiantil en actividades y programas para su aprendizaje y desarrollo personal.
- Los datos arrojados por estudios hechos a la comunidad estudiantil, a saber: encuestas estudiantiles, uso del tiempo del estudiantado.

En caso de que las autoridades académicas tengan disponibles los siguientes indicadores, también constituirían un insumo de gran interés para el análisis de datos educativos:

- Tasa de progreso en el aprendizaje de habilidades y competencias específicas.
- Tasa de retención.
- Tasa de graduación.
- Tasa de empleo de egresados.

Actualmente en la BDESSalud no se registran en Aleph las reservas o apartados de material bibliográfico. Sin embargo, el recopilar este tipo de solicitudes construiría una buena fuente de información para conocer cuáles son los materiales con mayor demanda, dado que se pueden identificar las características de las “filas de espera” que recrean en el sistema los usuarios para obtener el material de su interés en préstamo. En el caso de los préstamos interbibliotecarios no fueron incluidos en el estudio debido a que no son empleados en el SUBA-UACH. Sin embargo, podría considerarse su aplicación e inclusión en el DW en el caso de otras bibliotecas que sí apliquen tal servicio.

---

importante señalar que se trata de una correlación, no causalidad [...] La correlación entre dos variables no implica, por sí misma, ninguna relación de causalidad” (Villalobos y Portas, 2013, p. 7).

Otra posibilidad de análisis lo constituyen los metadatos derivados de la descripción bibliográfica del MDoc<sup>70</sup> que pueden ser obtenidos del SIGB Aleph, los cuales pueden ser empleados para delimitar interrelaciones temáticas entre los materiales para desarrollar sistemas de recomendación bibliográfica, así como estudios de uso de materiales con temáticas determinadas.

En cuanto a los datos referentes al tipo de transacciones (tanto en formato físico como en digital) efectuadas en la biblioteca, resultan necesarios para determinar los patrones de búsqueda que se dan en cuanto al uso del catálogo, los servicios de información y recursos electrónicos disponibles en el centro de información. Para esta tesis únicamente se contó con datos referentes a las transacciones en formato físico, pero una posibilidad excelente de análisis lo constituyen los datos de las interacciones de los usuarios con las bases de datos electrónicas, sobre todo considerando que, a raíz de la pandemia y la implementación de nuevas modalidades de acceso a la información de la BDESSalud, es muy probable que el uso de estos recursos haya tenido un aumento significativo.

En lo que respecta a la verificación de las hipótesis de investigación, es necesario recordar de qué manera fueron planteadas al inicio de esta tesis:

**HIPÓTESIS 1:** No existe una relación completamente funcional entre: 1) las estadísticas generadas por las actividades bibliotecarias y, 2) las necesidades mostradas en los procesos de evaluación y toma de decisiones de estas actividades.

**HIPÓTESIS 2:** La bibliominería tiene una relación directamente proporcional con la calidad de la gestión bibliotecaria, dado que contribuye a la optimización de los datos primarios y eleva su potencial de utilidad en los procesos de evaluación y toma de decisiones de las actividades bibliotecarias.

**HIPÓTESIS 3:** Considerando lo anterior, resultaría factible el diseño de un modelo de bibliominería que garantice las relaciones teóricas y metodológicas necesarias en la identificación de patrones de uso que generen conocimiento significativo orientado a la evaluación y a la toma de decisiones de actividades bibliotecarias.

---

<sup>70</sup> Véase: Retro y prospectiva de la investigación bibliotecológica sobre metadatos (Rodríguez, 2020).

Luego del trabajo realizado es posible decir que, dadas las limitaciones existentes entre los sistemas de gestión bibliotecaria tradicionales, no siempre es posible que éstos satisfagan completamente las necesidades de información que suponen la evaluación y toma de decisiones. La bibliominería puede constituir un aliado para afrontar esto, pues al formar parte de la metodología de KDD apoya en la labor de “transformar datos de bajo nivel en conocimiento de alto nivel” (Chang y Chen, 2006, p. 310), en el caso de esta tesis permitió tener un mayor conocimiento del comportamiento de uso de MDoc por parte de los cuatro perfiles de usuarios detectados, información que puede ser empleada por la jefatura de biblioteca para mejorar el diseño de la provisión de servicios en busca de garantizar la calidad de los mismos. Esta tesis permitió probar la verosimilitud de las tres hipótesis planteadas. Sin embargo, se considera que para mejorar las probabilidades de éxito al implementar un proyecto de bibliominería es necesario cumplir con una serie de condiciones o requisitos mínimos necesarios para que se asegure una adecuada obtención, tratamiento y uso de los datos. De acuerdo con la experiencia recogida en el desarrollo de esta investigación, tales condiciones o requisitos son:

- a. **DISPONIBILIDAD.** ¿La organización cuenta con los datos necesarios para emprender tareas de Ciencia de datos para resolver la(s) problemática(s) detectada(s)?<sup>71</sup>
- b. **ACCESIBILIDAD.** Si se tienen los datos necesarios ¿estarán disponibles para llevar a cabo su tratamiento y análisis? ¿se cuenta con la infraestructura necesaria para almacenarlos/consultarlos/descargarlos/transformarlos? Lo anterior requiere definir una ruta estratégica para asegurar un flujo de datos adecuado y que esté

---

<sup>71</sup> ¿Estamos recabando los datos correctos en las cantidades correctas antes de enfocarnos en extraer información útil de ellos? ¿Cómo identificar qué datos recopilar y conservar y qué no conservar? Las posibles respuestas a estas preguntas dependen de los objetivos y metas de los usuarios, así como de la aplicación correcta del proceso KDD (Uthurusamy, 1996, p. 567).

alineado con las problemáticas o temas que la Ciencia de datos está tratando de resolver en la organización<sup>72</sup>.

- c. **IMPLEMENTABILIDAD.** En este apartado haremos referencia a la aplicación de los modelos derivados del análisis de los datos de la organización. Se debe contar con garantías de que los procesos de la organización son lo suficientemente maduros para implementar y consumir los modelos de Ciencia de datos que eventualmente se crearán (Raina y Krishnamurthy, 2022, p. 29-31). Asimismo, es importante considerar las limitaciones espaciales y temporales para el ingreso de la información al DW.

Como es posible vislumbrar, es importante que la biblioteca cuente con una estrategia de datos<sup>73</sup> bien fundamentada que pueda apoyar y no obstaculizar en los análisis que se requieren llevar a cabo. Eso incluye el mejorar la captura de los datos, lo que implica dar una apropiada capacitación al personal en el uso estandarizado de los formatos de captura, de manera que se asegure la integridad de la información contenida en la base de datos.

## 4.2 Sobre los resultados y aportes del estudio

La principal aportación de esta tesis fue el ofrecer un panorama detallado de las implicaciones que conlleva el emplear elementos de Ciencia de datos en los espacios bibliotecarios, a partir del ejercicio de construcción del DW de la BDESSalud y el desarrollo del proceso de bibliominería efectuado. Las principales reflexiones derivadas de los resultados de esta tesis se muestran en los siguientes apartados:

---

<sup>72</sup> Véanse Raina y Krishnamurthy (2022, p. 30) y Peterman et al. (2021), quienes desarrollan aspectos teóricos y prácticos relacionados a la configuración de data lakes o “espacios dónde se recopilan los activos de datos” organizacionales de manera ordenada y estructurada para facilitar su posterior análisis.

<sup>73</sup> La estrategia de datos se refiere a la visión general en torno a la captura y utilización de los datos orientada a lograr los objetivos [organizacionales]. (Raina y Krishnamurthy, 2022, p. 309).

## **DW, BIBLIOMINERÍA Y USUARIOS DE LA INFORMACIÓN**

Como es notorio en el apartado de resultados de la tesis, al menos la mitad de las consultas planteadas como ejemplos de la funcionalidad del DW de la BDESSalud fueron efectuadas para conocer el comportamiento informativo de los usuarios de la biblioteca<sup>74</sup>. De acuerdo con Siguenza-Guzman (2015, p. 134), los principales aspectos de la biblioteca que son abordados a partir de [la aplicación de bibliominería] son la interacción de los usuarios con el sistema, la usabilidad de los sitios web de la biblioteca y la categorización de los usuarios considerando [tanto] el uso de la colección [como] la interacción de éstos con el sistema. Entre los resultados obtenidos resalta el que los usuarios más intensivos de los servicios de información son los estudiantes, quienes concentran la mayor proporción de transacciones en la biblioteca.

El ejercicio de bibliominería permitió identificar cuatro grupos de usuarios que constituyen los perfiles de los usuarios intensivos de los recursos ofrecidos por la BDESSalud. De acuerdo con Pal (2011, p. 12), los patrones de actividad descubiertos a partir de procesos de bibliominería pueden ofrecer beneficios en tres niveles diferentes:

- a. Beneficio para los usuarios a partir de la mejora de los servicios bibliotecarios
- b. Beneficios para la administración de la biblioteca a partir de la provisión de información mejorada para la toma de decisiones
- c. Beneficios para la institución a la cual pertenece y sirve la biblioteca, a través del informe de patrones relevantes de comportamientos de los usuarios.

La bibliominería de los datos de circulación puede proporcionar una serie de conocimientos sobre los grupos [de usuarios] que utilizan la biblioteca. Agrupando a los usuarios por los materiales [que hayan sido prestados] y vinculando la información demográfica a cada grupo, la biblioteca puede determinar grupos de usuarios conceptuales que proporcionan un modelo de los grupos importantes en la base de usuarios de la institución. Si la biblioteca ha [recopilado] datos de uso, se puede averiguar

---

<sup>74</sup> Las consultas 1, 5 y 6 involucraron a los usuarios del MDoc de la BDESSalud.

aún más sobre los patrones de uso [de información por parte] de los grupos de usuarios. El concepto de grupo de usuarios puede satisfacer algunas necesidades organizacionales comunes para comprender dónde residen los intereses y conocimientos comunes en la comunidad de usuarios. Al ayudar a identificar grupos de individuos con intereses comunes y/o experiencia en un tema, la bibliominería puede constituirse como un componente clave de los sistemas de gestión del conocimiento. Esta capacidad puede ser especialmente valiosa en grandes organizaciones donde los esfuerzos de investigación y desarrollo están dispersos en múltiples lugares (Nicholson y Stanton, 2003, p. 18).

En particular, los resultados del ejercicio de bibliominería en donde se describen las características de los cuatro grupos de usuarios intensivos identificados a partir de la aplicación del algoritmo *K-Means* ofrece un panorama que debe considerarse para moldear y mejorar los servicios y colecciones de la BDESSalud. Esta descripción es donde se ubican la mayoría de los usuarios reales del centro de información y, por ejemplo, dados los datos referentes a la antigüedad de la colección que consultan (que va en promedio, de 12.7 a 16.1 años, tomando como año base al 2019) se hace necesario el ahondar en las causas que motivan este comportamiento.

Como indican Chang y Chen (2006, p. 307), el objetivo principal de la minería de datos es mejorar la calidad de la interacción entre la biblioteca y sus usuarios. Los datos recabados contienen información valiosa que puede ser integrada en la estrategia de la biblioteca y puede ser usada para mejorar [el proceso de toma de decisiones].

## **DW, BIBLIOMINERÍA Y EL DESARROLLO DE COLECCIONES**

A partir de la aplicación de las metodologías empleadas para cumplir los objetivos de esta tesis (construcción del DW y el ejercicio de bibliominería) fue posible detectar cuáles son las áreas temáticas más demandadas por sus usuarios más intensivos. Esto fue con la ayuda de las consultas sobre títulos más consultados agrupados por tema, tipo de usuario y carrera (consulta 1), actualidad de la colección del MDoc por área temática (consulta 3) y tipo de devolución vencida por tipo de usuario, según el tema del MDoc



(consulta 6), así como del ejercicio de bibliominería que arrojó los temas más consultados por cada uno de los cuatro perfiles de usuarios detectados. Estas aproximaciones constituyen herramientas para lograr mantener actualizada la colección de la BDESSalud, para aprovecharlas es necesario que los bibliotecarios tengan bien identificados los rangos de clasificación del MDoc que corresponden a los temas de interés estar en posibilidades de monitorear su estado.

Otro de los ejercicios de utilidad para el área de Desarrollo de Colecciones lo constituye la consulta 4 (Tipo de adquisición del MDoc solicitado en préstamo por temporalidad). En particular, sus resultados permiten monitorear cuál es el tipo de adquisición de los materiales que están en circulación. Fue interesante ver que, a partir del 2016, aumenta el uso del MDoc donado por titulación. En el caso de la BDESSalud esto es de suma importancia, pues para este centro es frecuente la falta de presupuesto para la compra de materiales. Se considera que este reporte puede ser de utilidad para negociar con autoridades académicas la asignación presupuestal necesaria para fortalecer las colecciones.

Para Pavón, los beneficios de la bibliominería incluyen el “identificar quiénes son los usuarios reales que hacen uso de la colección y que son beneficiarios de las colecciones y servicios bibliotecarios, para adecuar a éstos tanto colecciones como servicios [...] conocer qué porcentaje de la colección es el que realmente está soportando la demanda de información y a partir de esto aplicar políticas que ayuden a reforzar las áreas más utilizadas de la colección, así como también detectar las áreas débiles y que requieren de mayor atención [...] corregir los libros con clasificaciones erróneas que anulan su localización en estantería, [etc.]” (Pavón, 2005, p. 97).

Con respecto a la planeación estratégica de las bibliotecas, la bibliominería puede apoyar en la generación de nuevas perspectivas estratégicas, por ejemplo, a partir del análisis del comportamiento de los usuarios en el proceso de circulación del material bibliográfico: es posible encontrar tendencias en cuanto al uso del MDoc y encontrar claves útiles para el desarrollo de colecciones más aptas para la población usuaria.

## **DW, BIBLIOMINERÍA Y EL DESEMPEÑO DEL PERSONAL BIBLIOTECARIO**

Las consultas que involucraron una revisión de las acciones del personal bibliotecario de la BDESSalud fueron la consulta 2 (Transacciones realizadas por bibliotecario) y la consulta 5 (Multas canceladas por bibliotecario, según el tipo de usuario y carrera), las cuales constituyen un complemento (sobre todo de visualización global) a los reportes disponibles en el LogAleph. De esta manera se da una muestra de cómo el DW puede brindar también herramientas para monitorear las actividades y desempeño de los empleados e identificar áreas de oportunidad en las actividades bibliotecarias.

### **4.3 Limitantes del estudio**

Durante el desarrollo de esta tesis hubo la necesidad de sobreponerse a ciertas limitantes que se presentaron, las cuales influyeron de manera determinante en la toma de decisiones sobre las acciones a llevar a cabo en las distintas etapas del proceso. A continuación, se hace un recuento de ellas:

#### **CARACTERÍSTICAS DE LOS DATOS DISPONIBLES**

Siguenza-Guzman expresó en su estudio (2015, p. 162) que la calidad y la eficacia de los resultados [obtenidos en los procesos de bibliominería] dependen en gran medida de la disponibilidad de la información relevante y de la calidad de los datos: cuantos más datos se proporcionen, se obtendrán resultados más precisos.

En el caso de este proyecto se contó con información proveniente de las tablas Oracle Aleph que mostraron inconsistencias y falta de información que se trató de subsanar lo más posible en el proceso ETL de la configuración del DW. Sin embargo, dadas las políticas de depuración de la base de datos transaccional en el rubro referente a usuarios y a la pérdida de información por fallas registradas en el servidor Aleph fue notoria la falta de datos importantes en el periodo temporal estudiado (2010-2019). Lo anterior implicó que se trabajara únicamente con los registros que estaban más completos.

Fue notoria también la existencia de datos incompletos o inexactos. Esto, de acuerdo con Gutiérrez y Barranco (2008, p. 91), puede ser atribuible a que los campos no se encuentran normalizados y el personal bibliotecario ingresa los datos al sistema “sin ninguna restricción”, lo que da pie a errores de escritura y a la inconsistencia de los datos.

Asimismo, fue posible verificar que no se emplea el campo designado por Aleph para la captura del año de publicación del material documental (Z13-YEAR), lo cual denota una subutilización del SIGB y, aparte de incrementar la carga para el personal de sistemas al requerir que realice una consulta *SQL* especial para obtener el dato de año de publicación mediante una sustracción del campo de clasificación, da pie a errores al obtener el dato debido a la variabilidad de la longitud de las cadenas de clasificación y a que la consulta llega a recuperar caracteres que no corresponden al dato solicitado.

## **PREPARACIÓN DEL PERSONAL INVOLUCRADO**

De acuerdo con Ball (2019, pp. 3 y 8), los principales factores limitantes para la realización de proyectos que involucran el análisis de datos en bibliotecas suelen ser la falta de personal calificado, las frecuentes carencias de infraestructura, los desafíos técnicos (entre los que se cuentan la diversidad de formatos de datos y las herramientas para su análisis), la protección de datos, problemas de financiamiento, incertidumbres legales, entre otros. Esto coincide con el planteamiento hecho por Nemati y Barko (2003, p. 291), quienes señalan que las organizaciones (especialmente las pequeñas) sin experiencia o competencias en el trabajo con datos, deben buscar apoyo fuera de su organización para la adecuada implementación de estos proyectos.

En lo personal, la autora de la tesis debió prepararse para contar con las bases necesarias para comprender las implicaciones (sobre todo teóricas y en cierta medida prácticas) del diseño e implementación de un DW y el procedimiento de bibliominería. Para ello tomó cursos relacionados a los temas de uso avanzado en Excel, introducción al diseño de bases de datos relacionales, *SQL* estándar básico y *Data warehouse* y minería de datos impartidos en la Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC) de la Universidad Nacional Autónoma de México

(UNAM). Empero, para la adecuada implementación del DW y la afinación del diseño y aplicación del ejercicio de bibliominería, buscó el apoyo de un Ingeniero en Sistemas del Instituto Politécnico Nacional (IPN) y de un Maestro en Ciencias de la Computación del Centro de Investigaciones en Matemáticas, A. C. (CIMAT). En este aspecto, se conformó un equipo para el desarrollo del proyecto de aplicación de la Ciencia de datos en la BDESSalud que obtuvo resultados satisfactorios en la implementación del ejercicio.

Es importante considerar que la exigencia actual de una toma de decisiones basada en evidencia requiere que las bibliotecas cuenten con personal capacitado para trabajar activamente con datos, profesionales capaces de planificar qué evaluaciones son prioritarias y cuáles son los datos más apropiados para llevarlas a cabo, de manera que se puedan realizar análisis que reflejen tanto el valor o buenas prácticas del centro de información bajo estudio, así como las áreas de oportunidad que deben ser atendidas (ACRL, 2017). Es claro que las organizaciones que planean usar [la bibliominería para mejorar sus actividades] pueden requerir una importante inversión en cuanto a infraestructura de tecnologías de información y contratación [o capacitación] de personal técnico, pero aquellas bibliotecas que son pequeñas y/o que no cuentan con suficiente apoyo presupuestario, pueden ver la manera de organizarse con otras instituciones similares y sumar esfuerzos para llevar a cabo proyectos de análisis conjunto de datos” (Chowdappa, p. 88).

## **ALTA ROTACIÓN DEL PERSONAL DE SISTEMAS DEL SUBA**

Fue notoria una alta movilidad laboral del personal asignado a la Unidad de Sistemas en el SUBA. Durante la realización de esta tesis se identificaron cuatro personas distintas que ocuparon la Jefatura de la Unidad de Sistemas y con las cuales se estableció contacto para obtener los datos necesarios para la realización del proyecto. Esto conllevó el reto de que, la falta de familiaridad con los sistemas en uso ralentizó las operaciones necesarias para obtener la información requerida, pues éstos (sobre todo al inicio de su gestión) representaban una caja negra para sus administradores. Hubo muchas

ocasiones en que también hubo apatía de los participantes, lo cual implicó también restricciones con respecto al acceso a las fuentes de datos.

## 4.4 Proyecciones para el estudio

Dentro de las posibles vertientes que pueden emanar del tema de investigación de esta tesis para investigaciones posteriores resaltan las siguientes:

- Averiguar cuáles han sido los cambios en los perfiles de puesto de los bibliotecarios asociados al uso y aplicación de metodologías de la Ciencia de datos.
- Delimitación de los posibles roles (y sus características) que pueden estar disponibles para los profesionales de la información referidos a la Ciencia de datos, como pueden ser:
  - Bibliotecario de sistemas (*Systems librarian*)<sup>75</sup>
  - Bibliotecario de datos (*Data librarian*)
  - Bibliotecario científico de datos (*Data scientist librarian*)
- Analizar si ha habido cambios en los programas de estudio del área Bibliotecológica y de la Información derivados de los nuevos requerimientos de formación académica y profesional para colegas involucrados en el uso intensivo de datos.
- Indagar cuáles deben ser las características del bibliotecario de datos.

Lewis y McGrath consideran el fenómeno de la “datificación” como resultado de una “(r)evolución” tanto de organizaciones como individuos para adaptarse, sobrevivir a y sacar provecho de la avalancha informativa que caracterizan los tiempos actuales. Empero, una consecuencia inesperada de estos cambios revolucionarios ha sido el surgimiento de una “brecha de talento”, término empleado para describir el desafío [...] de encontrar personas con las habilidades y preparación necesarias para extraer y

---

<sup>75</sup> Este perfil profesional es mencionado en seis ocasiones en el texto de las *Collected Oracle Tables* de la versión 23 del Aleph (Ex Libris, 2017).

analizar grandes cantidades de datos (estructurados y no estructurados) que permitan generar información significativa. De hecho, la demanda de esas habilidades se ha materializado tan rápidamente que, las fuentes tradicionales de suministro de nuevos talentos (colegios y universidades) han estado mal equipados para desarrollarlos a la escala y ritmo exigidos (Lewis y MacGrat, 2019, p. 2).

Los profesionales de la información cuentan con una larga tradición y, por ende, experiencia en el manejo de datos, sobre todo en las tareas de administración y organización de la información. Lo anterior constituye una habilidad central en la Ciencia de datos, que se manifiesta con mayor fuerza en el área de curaduría de datos. Muchos bibliotecarios también son excelentes comunicadores que han sido capacitados en el arte y la ciencia de transformar las necesidades de información del usuario en estrategias y recursos para la investigación y el aprendizaje (Stanton, 2012). Con esto se reconoce que dentro de su formación se pueden rescatar habilidades y conocimientos técnicos y contextuales de real importancia para la administración de datos. Empero, el panorama actual les demanda seguirse preparando para afrontar los retos que supone el abordar el corazón de la ciencia de los datos. Para adaptarse eficientemente al nuevo entorno, el bibliotecario académico deberá solventar al menos dos importantes obstáculos:

- **Brecha de habilidades:** Muchos bibliotecarios carecen de las habilidades técnicas necesarias para desenvolverse efectivamente en un ambiente de investigación rico en datos (Burton et al, 2018, p. 4). El procesamiento de datos puede presentar desafíos técnicos importantes, pues implica habilidades técnicas como la captura, almacenamiento, procesamiento y presentación [o visualización] de datos. Los datos que maneja una biblioteca típica son de diferentes tipos y pueden estar contenidos en fuentes de información heterogéneas. Debido a la variedad de tipos y formatos de información, integrarlos se convierte en una tarea compleja (Chowdappa, 2021, p. 88). Esto puede subsanarse a partir de la incorporación de los interesados en programas de capacitación que les permitan

obtener nuevos conocimientos y desarrollar las habilidades técnicas necesarias para el correcto manejo de los datos<sup>76</sup>.

- **Brecha administrativa:** Poca o nula “capacidad de los administradores de la biblioteca para comprender y valorar los beneficios de las habilidades internas en ciencias de datos y para proporcionar soporte organizativo y gerencial” (Burton et al, 2018, p. 5). Los directivos necesitan conocimientos, herramientas y sistemas que les permitan reconocer la trascendencia del empleo estratégico de la Ciencia de datos, de manera que la apliquen para la toma de decisiones y la gestión de las operaciones bibliotecarias.

Es necesario resaltar que “un bibliotecario de datos no necesita convertirse en programador, estadístico o administrador de bases de datos, pero debería estar interesado en aprender sobre los lenguajes y lógica de programación de computadoras, bases datos y herramientas de recuperación de información”. (Ribas et al, 2017, p.1). Se debe considerar que la falta de personal con una adecuada preparación teórica y sobre todo técnica para el análisis de datos puede ser un importante obstáculo para realizar aplicaciones de Ciencia de datos en centros de información. El elemento clave es que los analistas de datos no sólo necesitan habilidades estadísticas y de computación, sino también habilidades de conocimiento de dominio y capacidad de colaboración. La mayor parte de los profesionales de la información carecen de habilidades [teóricas y] técnicas, por lo cual enfrentan muchas dificultades para procesar y administrar [el análisis de datos]. Es necesario formar a estos profesionales en los aspectos relacionados a [la aplicación de Ciencia de datos en centros de información] y parece que el entrenamiento en formato de cursos cortos puede no ser suficiente (Chowdappa, 2021, pp. 87-88).

Así, una de las principales proyecciones de los planteamientos de esta tesis es el indagar si se han suscitado cambios importantes con respecto a los perfiles de puestos de bibliotecarios asociados al uso y aplicación de metodologías de la Ciencia de datos. En

---

<sup>76</sup> Véase Anexo 11. Temáticas para considerar en la formación de los bibliotecarios de datos.

los informes de tendencias de la Asociación de Bibliotecas Universitarias y de Investigación de Estados Unidos (ACRL, por las siglas en inglés de *Association of College and Research Libraries*) correspondientes a los años 2016, 2017, 2018 y 2020 se hace hincapié en la importancia de que los profesionales de la información estén preparados para involucrarse y participar activamente en proyectos que supongan el uso de técnicas y herramientas avanzadas de análisis de datos que sean de utilidad en las tareas de evaluación y análisis de colecciones y servicios bibliotecarios, así como para ofrecer servicios de administración de datos de investigación a los usuarios que así lo requieran.

En este mismo tono se leen los reportes de tendencias publicados por la IFLA en 2017 y 2022 (*International Federation of Library Associations and Institutions*, 2017 y 2022), en los cuales se hace hincapié en lograr que los profesionales de la información tengan la alfabetización en datos apropiada, pues identifican a los datos como un elemento clave que ayuda a generar e impulsar servicios más personalizados, permite analizar el comportamiento pasado [de los usuarios], hacer predicciones a futuro y responder a [sus] necesidades, además de que su uso apropiado fortalece la evaluación y la valoración del impacto de los servicios bibliotecarios.

Dentro de las principales ocupaciones de los administradores bibliotecarios se encuentra el lograr el equilibrio entre garantizar el funcionamiento de las operaciones necesarias para mantener un nivel de servicios adecuado y el hacer frente a nuevas tendencias que den paso a la innovación y hagan más atractiva la biblioteca para sus usuarios. Muchos podrán suponer que la búsqueda por adoptar conocimientos y habilidades en el área de la Ciencia de los datos puede llegar a interferir con la atención brindada a las áreas tradicionales del conocimiento bibliotecológico. Sin embargo, puede resultar de gran ayuda en la optimización de las tareas ya existentes, pues a partir de aplicaciones como los DW y la bibliominería, se tiene la posibilidad de emplear tales conocimientos de manera proactiva, analizar la información acumulada para encontrar patrones de relación y tendencias, de manera que se refuerce a la evaluación y la toma de decisiones en las áreas de interés de los administradores. Es en éste último rubro que el análisis de datos cobra una importancia central: lo idóneo es contar siempre con evidencia objetiva,



fundamentada en los hechos para tener acceso a una toma de decisiones basada en la evidencia (lo anterior sin menosprecio de los esfuerzos cualitativos derivados de actividades como entrevistas, descripciones y observaciones, recolección de evidencia anecdótica, desarrollo de grupos focales, etc., que sirven para dar contexto y mejorar la interpretación del análisis de los datos).

Otra vertiente para la continuación del estudio planteado en este documento es el continuar desarrollando la Interfaz de consulta y visualización del DW de la BDESSalud para ver la pertinencia de la inclusión de nuevas consultas de interés, así como el buscar la validación tanto de la interfaz de usuario (UI) como la experiencia del usuario (UX). De esta manera se esperaría que el prototipo de la interfaz de consulta y visualización integre de manera eficaz los elementos que comprende una buena experiencia del usuario de acuerdo con el modelo Honeycomb, a saber, que la interfaz sea:

1. Útil
2. Utilizable (usable)
3. Valiosa
4. Deseable
5. Localizable (disponible)
6. Creíble
7. Accesible (Kim, 2020, p. 229).

# **CAPÍTULO 5. Conclusiones y recomendaciones**

Los resultados que fueron obtenidos al desarrollar esta tesis pueden ser útiles para los profesionales de la información de bibliotecas de corte académico que estén interesados en realizar aplicaciones de bibliominería y DW como parte de un esquema que aporte a las actividades de evaluación y toma de decisiones. Las principales aportaciones de esta investigación son:

1. El brindar los detalles de cómo fue estructurado el DW, el cual constituyó una estructura de datos que permitió aprovechar éstos al máximo en el momento de realizar el ejercicio de bibliominería.
2. A partir del ejercicio de bibliominería, fue posible identificar cuáles son los patrones de uso presentados por quienes emplean de manera intensiva el servicio de préstamo de materiales documentales de la BDESSalud.
3. La estructuración del DW y la aplicación del ejercicio de bibliominería permitió establecer un modelo de bibliominería cuyas características, ajustadas a los requerimientos de la BDESSalud, permitió brindarles a los tomadores de decisiones información relevante o significativa para la toma de decisiones en su centro de trabajo.
4. Asimismo, otra aportación de esta tesis es la actualización del estado del arte sobre las aplicaciones prácticas de Ciencia de datos en el área bibliotecológica en México, pues durante su desarrollo se procuró llevar a cabo la identificación o mapeo de las aproximaciones en México en el uso de la Ciencia de datos aplicado a los procesos bibliotecarios. De esta manera se identificaron ciertos acercamientos que se enlistan en el Anexo 12. Identificación de iniciativas de aplicación de metodologías de la Ciencia de datos en bibliotecas mexicanas.

Este apartado muestra cuáles fueron las conclusiones derivadas de los resultados de investigación de mayor interés de acuerdo con los planteamientos iniciales del proyecto,

asimismo se ofrecen recomendaciones que pueden permitir darle un adecuado seguimiento al tema para fincar su desarrollo futuro.

## 5.1 Conclusiones

El objetivo primario de la tesis fue el delimitar las estructuras de datos, metodologías y herramientas que propiciaran la identificación de patrones de uso que generen conocimiento significativo con el propósito de establecer su relación con los procesos de evaluación y la toma de decisiones en las actividades bibliotecarias. Esto se logró a partir del diseño e implementación del DW de la BDESSalud empleando el programa *MySQL Workbench* y la metodología Hefesto, cuyas principales ventajas son un mayor dinamismo en la presentación de datos que son fáciles de comprender, verificables y que constituyen una base sobre la cual se pueden desarrollar futuros análisis que dependerán de las necesidades de los tomadores de decisiones de la BDESSalud.

Durante el desarrollo del DW fue posible constatar que el tomar el tiempo necesario para definir las necesidades y problemáticas a que se enfrentan los tomadores de decisiones es un paso fundamental para el adecuado desarrollo de un proyecto de DW y bibliominería. Una apropiada definición de dichas necesidades contribuirá también a identificar cuáles son los datos que aportarán mayores beneficios al ser incluidos en el DW, pues de esta manera será posible establecer consultas pertinentes. En el caso de esta tesis, las consultas derivadas del DW permiten obtener reportes útiles para el desarrollo de colecciones, la supervisión y evaluación del desempeño de los bibliotecarios, la actualización del MDoc por área temática, el tipo de adquisición de MDoc predominante, así como el tipo de devolución vencida según el tipo de usuario de acuerdo con tema del MDoc. Todas estas consultas son complementarias al ejercicio de bibliominería propuesto.

Por su parte, el ejercicio de bibliominería permitió identificar cuatro grupos de usuarios que constituyen los perfiles de los usuarios intensivos de los recursos ofrecidos por la BDESSalud. Un hallazgo relevante es que el 67% de los usuarios intensivos son mujeres.

El **grupo 1** está constituido completamente por mujeres, de las cuales el 84% pertenece a la carrera de Médico Cirujano Partero y pertenecen a la Facultad de Medicina y Ciencias Biomédicas. Es el grupo con mayor número de usuarios de Posgrado y Académicos. Tienen el mayor promedio de préstamos mensuales (2) y consultan el material más actualizado dentro de los 4 grupos identificados (con una antigüedad promedio de 12.7 años). Consultan los temas de *“Human anatomy, cytology, histology”*; *“diseases”*; *“biochemistry”*; *“knowledge”* y *“Physiology and related subjects”*.

El **grupo 2** se conforma en un 78% por mujeres. El 81% de los usuarios en este grupo pertenece a la Licenciatura en Enfermería, de la Facultad de Enfermería y Nutriología. Junto con el grupo 4, manejan un promedio mensual de 1.4 libros. Este segmento consulta el MDoc más antiguo entre los grupos identificados (16.1 años de antigüedad promedio) y consulta una menor diversidad de temas que el resto. Entre los temas destacados para este grupo encontramos *“Human physiology”*, *“Personal health and safety”*, *“Food and drink”*, *“Chemistry and allied sciences”* y *“Forensic medicine; incidence of injuries, wounds, disease; public preventive medicine”*.

El **grupo 3** se integra completamente por hombres, de los cuales el 83% pertenece a la carrera de Médico Cirujano Partero de la Facultad de Medicina y Ciencias Biomédicas. Manejan el segundo mejor promedio mensual de préstamos de 1.9 libros, una décima abajo del grupo 1. Similar a este, el grupo 3 también consulta materiales actualizados (con promedio de 13.1 años de antigüedad). Este grupo, al igual que el Grupo 1, recurren continuamente al material de *“Human anatomy, cytology, histology”*; *“Diseases”*; *“Biochemistry”* y *“Physiology and related subjects”*. Empero, el detalle con este grupo es que presenta el mayor promedio de días de retraso por préstamo, lo cual indica que, de prestarse MDoc a algún usuario con estas características, la probabilidad de que devuelva tarde el MDoc aumenta.

El **grupo 4** está formado en un 80% por mujeres. El 94% de sus integrantes pertenece a la Licenciatura en Nutrición de la Facultad de Enfermería y Nutriología. Maneja un promedio mensual de préstamos igual al grupo 2 (1.4 libros), y consultan materiales con una antigüedad promedio de 14.8 años. El grupo se destaca por consultar los temas de *“Human physiology”*, *“Personal health and safety”*, *“Food and drink”*, *“Chemistry and*

*allied sciences*” y “*Forensic medicine; incidence of injuries, wounds, disease; public preventive medicine*”.

Durante la revisión de las salidas de la interfaz para consulta y visualización del DW de la BDESSalud fue posible dilucidar que las salidas por sí mismas no darían el *insight* necesario para resolver o comprender una problemática: fue necesario emplear esos insumos para indagar las dudas que surgían con respecto a las consultas realizadas. Por ello es posible decir que el conocimiento significativo es de quien lo trabaja: fue necesario utilizar esas salidas y reprocesarlas (sobre todo usando Excel, lo cual ocurrió para las salidas de las consultas 3 y 4) para tener nociones más claras de lo que estaba sucediendo. Empero, esto no quita el valor agregado que las mismas salidas ofrecieron: los resultados obtenidos no habrían sido posibles si únicamente se hubieran empleado los insumos tradicionales presentes en la biblioteca (las consultas directas al Aleph y los reportes LogAleph).

En sí, la obtención de los perfiles de usuarios a partir del ejercicio de bibliominería constituyen un insumo importante para la toma de decisiones en la BDESSalud. Sin embargo, será a partir del análisis de los patrones de uso del MDoc que presentan tales perfiles que se podrá obtener conocimiento valioso para poder instrumentar acciones significativas en el sistema. La experiencia en la aplicación de la bibliominería permitió verificar que la MD no es un proceso que en automático da acceso a un conocimiento profundo sobre el proceso estudiado. Pero entonces ¿de qué depende el conocimiento significativo? Podemos decir que es una suma de factores que incluye:

- Uso de herramientas para el análisis avanzado de los datos (DW y bibliominería).
- Resultados o productos derivados de la aplicación de las herramientas mencionadas.
- Juicio y experiencia de quien toma las decisiones al analizar los resultados.
- Espíritu inquisitivo del decisor.
- Intención estratégica de los decisores para aplicar acciones con base en los resultados obtenidos.
- Compromiso de los involucrados en el proceso de evaluación y toma de decisiones.

Como puede verse, no es por arte de magia que se obtengan resultados satisfactorios. El proceso de desarrollo del DW y de la aplicación de la bibliominería consiste en aproximaciones interactivas, iterativas y reiterativas. Es un continuum de experimentación y mejora que implica la necesidad de “arrastrar el lápiz” y tener los objetivos muy claros con respecto a lo que se busca.

Por ejemplo, en el caso de los resultados obtenidos en la aplicación de la bibliominería, pueden surgir dudas con respecto a los usuarios de carreras que no se vieron representados en los grupos definidos por *K-Means*. Implica que se averigüe qué es lo que pasa con los usuarios de las carreras de la FEN y la FMCB que no lograron ser identificados como usuarios asiduos: ¿A qué puede deberse su falta de presencia? ¿Es por la temporalidad en que fueron creadas las carreras? ¿Existe un menor número de usuarios pertenecientes a esos programas de estudio? ¿O es simplemente desinterés en el uso de los servicios bibliotecarios por este segmento de usuarios? Dependerá de los decisores las vías a tomar para resolver tales interrogantes, en caso de que les sean relevantes.

Los resultados obtenidos en la aplicación de la bibliominería en el DW de la BDESSalud pueden emplearse para negociar con las autoridades el establecimiento de un presupuesto operativo que permita brindar servicios de calidad en la biblioteca, además de ser de utilidad como apoyo para la delimitación de políticas de préstamo apropiadas; mejorar la comunicación con la comunidad usuaria y, de acuerdo con los perfiles identificados, mantener a los usuarios informados con respecto a la llegada de materiales actualizados o novedosos que pertenezcan a sus áreas temáticas de interés. Esto es de suma importancia debido a que, gracias al ejercicio de bibliominería realizado, fue posible constatar que la consulta de material novedoso es más bien escasa, lo cual también puede decirse del préstamo de MDoc en general (el promedio mensual de préstamos no supera los dos libros). Esta segmentación de usuarios de acuerdo con sus patrones de uso del MDoc debe usarse en aras de mejorar la calidad de los servicios ofrecidos y brindarles beneficios a estos usuarios, lo cual incidiría en una mejora tanto de las capacidades gerenciales de los encargados del centro de información como de la calidad educativa

de la institución de la que éste forma parte. Para hacer patentes estos beneficios en la BDESSalud, es necesario que se consideren los resultados obtenidos a partir de la bibliominería y se tomen las medidas pertinentes para incentivar, por ejemplo, un mayor número de préstamos de MDoc y la consulta de materiales de mayor actualidad entre los distintos grupos usuarios. En sí, la obtención de los perfiles de usuarios a partir del ejercicio de bibliominería constituye un insumo importante para la toma de decisiones en la BDESSalud. Sin embargo, será a partir del análisis de los patrones de uso del MDoc que presentan tales perfiles que se podrá obtener conocimiento valioso para poder instrumentar acciones significativas en el sistema.

## 5.2 Recomendaciones

Entre las recomendaciones a considerar derivadas de esta tesis resalta el que, antes de involucrarse en las aplicaciones de bibliominería, es necesario que los centros de información cumplan con los requisitos de **disponibilidad** de los datos necesarios; **accesibilidad** a los mismos, y la garantía de que los resultados de los análisis sean **implementados** de manera satisfactoria en la organización sede.

Lo anterior implica que las organizaciones interesadas tengan desplegadas estrategias de gobernanza de datos que les oriente en la toma de acciones para mejorar la calidad de los datos. Por ejemplo, desde su captura: empleando una terminología estándar o previamente acordada. Eso puede ahorrar tiempo valioso en el proceso de limpieza de los datos y en su procesamiento previo a su análisis.

Con respecto al uso del DW creado para la BDESSalud, es necesario indicar que, dadas sus características, es susceptible de seguir procesos de mejora y crecimiento a futuro "alimentándose" con datos (tanto los preestablecidos como nuevos) proporcionados por la DESSalud, el SUBA y por otras áreas de la UACH. Por ejemplo, para retomar su proceso de ETL puede considerarse el uso de las distintas interfaces de programación de aplicaciones (API's) ofrecidas por ExLibris para el sistema Aleph. También es

importante el buscar la mejora de la interfaz de consulta y visualización del DW de la BDESSalud, de manera que se consoliden unas adecuadas características de UI y la experiencia que tienen los usuarios al emplearla.

Es necesario que los profesionales de la información cuenten con vías de preparación eficaces para emprender proyectos relacionados a la aplicación de elementos de la Ciencia de datos, como lo son la bibliominería y el diseño de DW. Esto puede ser ya sea a partir de materias enfocadas al análisis y aprovechamiento de datos dentro de los programas de estudio o de cursos profesionales fuera del currículo dedicados al tema. En este sentido es necesario también analizar si ha habido cambios en los programas de estudio del área Bibliotecológica y de la Información derivados de los nuevos requerimientos de formación académica y profesional para colegas involucrados en el uso intensivo de datos.

Otra de las aplicaciones de la bibliominería en centros de información es ofrecer el sustento fáctico necesario que sirva como apoyo en la rendición de cuentas. Generalmente las bibliotecas se encuentran bajo un fuerte escrutinio con respecto a su funcionamiento. No son poco frecuentes los cortes de presupuesto. De hecho, entre los problemas más comunes que atañen a las bibliotecas académicas se encuentran las limitaciones financieras que pueden suponer un obstáculo difícil de sortear a la hora de ofrecer servicios y productos informativos de calidad. Los reportes analíticos de datos fundamentados en la bibliominería pueden servir para demostrar la valía de las bibliotecas, de sus recursos y servicios. Priscille Dando invita a los bibliotecarios a tener a “los datos como el corazón de la estrategia de defensa” de sus centros de trabajo (Dando, 2014, p. viii). Indudablemente el uso adecuado de los datos referentes a la operación de nuestros centros de información nos brindará las armas necesarias para emprender estas tareas.



# Referencias

- Aceves Cano, Petra Pomposa, et al. (2002). *Historia de las bibliotecas de la Universidad Autónoma de Chihuahua*. Chihuahua, Chih.: Coordinación del Sistema Universitario de Bibliotecas Académicas, Dirección Académica, Universidad Autónoma de Chihuahua.
- Ahmad, Norita, Areeba Hamid y Vian Ahmed. (2022). Data science: hype and reality. *Computer*, 55, 95-101.
- Al-Barashdi, Hafidha y Rahma Al-Karousi. (2018). Big data in academic libraries: literature review and future research directions. *Journal of Information Studies and Technology*, 2(13). <http://www.qscience.com/doi/pdf/10.5339/jist.2018.13>
- American Library Association. (2014). *Data everywhere*. Chicago, Illinois: ALA. <http://www.ala.org/tools/future/trends/data>
- Association of College and Research Libraries, Research Planning and Review Committee. (2016). 2016 top trends in academic libraries: a review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News*, 77(6), 274-281.
- Association of College and Research Libraries, Research Planning and Review Committee. (2020). 2020 top trends in academic libraries: a review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News*, 81(6), p. 270. <https://doi.org/10.5860/crln.81.6.270>
- Association of College and Research Libraries, Research Planning and Review Committee. (2017). *Environmental scan 2017*. Chicago, Illinois: ACRL. <http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/EnvironmentalScan2017.pdf>
- Association of College and Research Libraries, Research Planning and Review Committee. (2018). 2018 top trends in academic libraries: a review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News*, 79(6), 286-300.
- Ball, Rafael. (2019). Big data and their impact on libraries. *American Journal of Information Science and Technology*, 3(1), 1-9.
- Banerjee, Kyle. (1998). Is data mining right for your library? *Computers in Libraries*, 18(10), 28-31.
- Bernabeu R. Dario y Mariano García Mattío. (2017). *Hefesto data warehousing: guía completa de aplicación teórico-práctica; metodología data warehouse*. Documento protegido con licencia Creative Commons BY-NC-ND 4.0 International. <http://www.redopenbi.com/profiles/blogs/libro-hefesto-data-warehousing-v3-pdf-y-online>

- Bollen, Johan, Rick Luce. (2002). Evaluation of digital library impact and user communities by analysis of usage patterns. *D-Lib Magazine*, 8(6). <http://www.dlib.org/dlib/june02/bollen/06bollen.html>
- Breeding, Marshall. (mayo, 2021). Informe sobre sistemas bibliotecarios 2021: avance de las tecnologías bibliotecarias en tiempos difíciles. *American Libraries*. <https://librarytechnology.org/document/26253>
- Breeding, Mashall. (mayo, 2020). Informe sobre sistemas de biblioteca 2020: nuevas oportunidades en medio de la consolidación. *American Libraries*. <https://librarytechnology.org/document/25247>
- Britannica, T. Editors of Encyclopaedia. (2021). *Pattern recognition*. *Encyclopedia Britannica*. <https://www.britannica.com/technology/pattern-recognition-computer-science>
- Budd, John M. (1998). *The academic library: its context, its purpose, and its operation*. Library and Information Science Text Series. Englewood, Colorado: Libraries Unlimited.
- Burton, Matt y Liz Lyon. (2017, abril/mayo). Data science in libraries. *Bulletin of the Association for Information Science and Technology*, 43(4), 33-35. <https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/bul2.2017.1720430409>
- Burton, Matt, Liz Lyon, Chris Erdmann y Bonnie Tijerina. (2018). *Shifting to data savvy: the future of Data Science in libraries* [reporte de proyecto]. Pittsburgh, Philadelphia: University of Pittsburgh, 24 p. <http://d-scholarship.pitt.edu/33891/1/Shifting%20to%20Data%20Savvy.pdf>
- Candás Romero, J. (2006). Minería de datos en bibliotecas: Bibliominería. [Versión electrónica], *BiD: Textos universitaris de biblioteconomia i documentació*, 17. <http://eprints.rclis.org/8812/>
- Chang, Chan-Chine y Ruey-Shun Chen. (2006). Using data mining technology to solve classification problems: a case study of campus digital library. *The electronic library*, 24(3). 307-321.
- Cheung, David; Williams, Graham J. y Li Qing, Editores. (2001). *Advances in knowledge discovery and data mining: 5<sup>th</sup> Pacific Asia conference; proceedings / PAKDD 2001, Hong Kong, China, April 16-18, 2001*. Berlín, Alemania: Springer.
- Chowdappa, V. (2021). Big data: applications and issues in library and information centres. *Research journey*, 279(A), 85-89. <https://www.shriwaghmarebrothers.com/wp-content/uploads/2021/12/Special-Issue-279-A-Rahata-College.pdf#page=86>
- Clarivate. (2021, 1º de diciembre). *Clarivate successfully completes acquisition of ProQuest* [comunicado de prensa]. <https://clarivate.com/news/clarivate-successfully-completes-acquisition-of-proquest/>

- Colle, Raymond. (2002). *Explotar la información noticiosa: data mining aplicado a la documentación periodística*. Madrid: Universidad Complutense de Madrid, Departamento de Biblioteconomía y Documentación.
- Consejo Nacional para Asuntos Bibliotecarios de las Instituciones de Educación Superior. (2012). *Normas para bibliotecas de instituciones de educación superior e investigación*. 2 ed. Colección Normatividad. La Paz, Baja California Sur: Comité Técnico para el Análisis y Actualización de las Normas del CONPAB-IES.
- Contreras Barrera, Marcial. (2014). Minería de texto: una visión actual. *Biblioteca universitaria*, 17(2), 129-138.
- Contreras Barrera, Marcial. (2018). *Minería de texto en la organización de la información documental* [Tesis de Doctorado, Universidad Nacional Autónoma de México]. TESIUNAM, Dirección General de Bibliotecas y Servicios Digitales de Información.
- Cox, Brian L. y Margie Jantti. (2012). Capturing business intelligence required for targeted marketing, demonstrating value, and driving process improvement. *Library and Information Science Research*, 34, 308-316.
- DAMA International. (2017). *DAMA-DMBOK: data management body of knowledge*. 2 ed. Basking Ridge, Nueva Jersey: Technics Publications.
- Dando, Priscille. (2014). *Say it with data: a concise guide to making your case and getting results*. Chicago: ALA Editions, American Library Association.
- Dekker, Harrison y Paula Lackie. (2016). Technical data skills for reproducible research. En Linda Kellam y Kristi Thompson (eds.), *Databrarianship* (pp. 93-112). Chicago, Ill.: Association of College and Research Libraries.
- Ex Libris. (2017). *ExLibris Aleph: collected Oracle tables, version 23*. SL: ExLibris Limited.
- Facultad de Medicina y Ciencias Biomédicas. (2021a). *Especialidades*. Chihuahua, Chih.: Facultad de Medicina y Ciencias Biomédicas, Universidad Autónoma de Chihuahua. <http://fm.uach.mx/conocenos/2011/09/14/especialidades/>
- Facultad de Medicina y Ciencias Biomédicas. (2021b). *Facultad de Medicina de la Universidad Autónoma de Chihuahua*. Chihuahua, Chih.: Facultad de Medicina y Ciencias Biomédicas, Universidad Autónoma de Chihuahua. <http://fm.uach.mx/>
- Facultad de Medicina y Ciencias Biomédicas. (2021c). *Maestrías de posgrado*. Chihuahua, Chih.: Facultad de Medicina y Ciencias Biomédicas, Universidad Autónoma de Chihuahua. [http://fm.uach.mx/portal/2018/11/21/\\_master/](http://fm.uach.mx/portal/2018/11/21/_master/)
- Fayyad, Usama M., Gregory Piatetsky-Shapiro, Padhraic Smyth y Ramasamy Uthurasamy, Editores. (1996). *Advances in knowledge discovery and data mining*. Menlo Park, California: American Association for Artificial Intelligence, The MIT Press.
- Fernández Morales, Mynor y Roger Bonilla Carrión. (2020). Bibliominería, datos y el proceso de toma de decisiones. *Revista Interamericana de Bibliotecología*, 43(2). <https://doi.org/10.17533/udea.rib.v43n2e18>

- Fisher, Shelagh y Tony Oulton. (1995). Information for decision making: the DECIMAL project. *Library review*, 44(8), 10-19.
- Garza Mercado, Ario. (2003). *Función y forma de la biblioteca universitaria: elementos de planeación administrativa para el diseño arquitectónico*. 2 ed. Jornadas, 83. México, D. F.: Biblioteca Daniel Cosío Villegas, El Colegio de México.
- Gorbea Portal, Salvador. (1994). Principios teóricos y metodológicos de los estudios métricos de la información. *Investigación bibliotecológica: archivonomía, bibliotecología e información*, 9(17).
- Gorbea Portal, Salvador. (1998). Modelación matemática de la actividad bibliotecaria: una revisión. *Investigación bibliotecológica: archivonomía, bibliotecología e información*, 12(24).
- Gorbea Portal, Salvador. (2013). Tendencias transdisciplinarias en los Estudios Métricos de la Información y su relación con la gestión de la información y del conocimiento. *Perspectivas em Gestão & Conhecimento*. 3(1), 13-27.
- Gorbea Portal, Salvador. (2016). Una nueva perspectiva teórica de la bibliometría basada en su dimensión histórica y sus referentes temporales. *Investigación bibliotecológica: archivonomía, bibliotecología e información*, 30(70), 11-16. <http://dx.doi.org/10.1016/j.ibbai.2016.10.001>
- Gorbea Portal, Salvador y María de Jesús Madera Jaramillo. (2017). Diseño de un data warehouse para medir el desarrollo disciplinar en instituciones académicas. *Investigación bibliotecológica: archivonomía, bibliotecología e información*, 31(72), 161-181. <http://dx.doi.org/10.22201/iibi.0187358xp.2017.72.57828>
- Grupo Sistemas Lógicos. (2017). *Aleph 500: ExLibris Aleph*. México, D. F.: GSL México. <http://www.gsl.com.mx/aleph-500/>
- Guenther, Kim. (2000). Applying data mining principles to a library data collection. *Computers in libraries*, 20(4), 60-63.
- Gutiérrez Hernández, Guadalupe Vanessa Carolina y Verónica Barranco Serrano. (2008). *Minería de datos dentro del proceso de KDD aplicado a la base de datos de circulación bibliográfica de la Biblioteca Central* [Tesis de Licenciatura, Universidad Nacional Autónoma de México]. TESIUNAM, Dirección General de Bibliotecas y Servicios Digitales de Información.
- Han, Jiawei, Micheline Kamber y Jian Pei. (2012). *Data mining: concepts and techniques*. Waltham, MA: Morgan Kaufmann Publishers.
- Haro Valle, Valeria Alexandra, Wilson Rodrigo Pérez Rocano, Lorena Siguenza-Guzman, Dirk Cattrysse y Víctor Hugo Saquicela Galarza. (2014). Diseño e implementación de un sistema de soporte de decisiones para el Centro de Documentación "Juan Bautista Vázquez". *MASKANA*, 5 (número especial: Actas del II Congreso Ecuatoriano de Tecnologías de la Información y Comunicación TIC.EC 2014), 245-256. <http://dspace.ucuenca.edu.ec/handle/123456789/21411>

- Haro Valle, Valeria y Wilson Rodrigo Pérez Rocano. (2014). *Data warehouse para el Centro de Documentación Regional "Juan Bautista Vázquez"*. [Tesis de Licenciatura, Universidad de Cuenca]. Repositorio Institucional Universidad de Cuenca.
- Hernández Orallo, José y Cèsar Ferri Ramírez. (2006). *Extracción automática de conocimiento en bases de datos e ingeniería del software. T. 2 Minería de datos y extracción de conocimiento de bases de datos. Programación declarativa e ingeniería de la programación*. Valencia, España: Dep. de Sistemes Informàtics i Computació. Universitat Politècnica de València. <http://users.dsic.upv.es/~jorallo/docent/doctorat/t2a.pdf>
- Hood, William W. y Concepción S. Wilson. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, 52(2), 291-314.
- Inmon, William Harvey. (2003). *Building the data warehouse*. 3 ed. Nueva York: John Wiley & Sons.
- International Federation of Library Associations and Institutions. (2013) *IFLA statement on text and data mining*. [S. l.]: International Federation of Library Associations and Institutions. <https://www.ifla.org/publications/node/8225>
- International Federation of Library Associations and Institutions. (2015). *IFLA statement on privacy in the library environment*. [S. l.]: International Federation of Library Associations and Institutions. <https://www.ifla.org/wp-content/uploads/2019/05/assets/hq/news/documents/ifla-statement-on-privacy-in-the-library-environment.pdf>
- International Federation of Library Associations and Institutions. (2017). *Riding the wave or caught in the tide? Navigating the evolving information environment*. Insights from the IFLA Trend Report. [https://trends.ifla.org/files/trends/assets/insights-from-the-ifla-trend-report\\_v3.pdf](https://trends.ifla.org/files/trends/assets/insights-from-the-ifla-trend-report_v3.pdf)
- International Federation of Library Associations and Institutions. (2020). *IFLA statement on libraries and artificial intelligence*. [S. l.]: International Federation of Library Associations and Institutions. [https://www.ifla.org/wp-content/uploads/2019/05/assets/faife/ifla\\_statement\\_on\\_libraries\\_and\\_artificial\\_intelligence.pdf](https://www.ifla.org/wp-content/uploads/2019/05/assets/faife/ifla_statement_on_libraries_and_artificial_intelligence.pdf)
- International Federation of Library Associations and Institutions. (2022). *IFLA trend report 2021 update: 20 political, economic, social, cultural and technological trends to shape the future of our field and the communities we serve, as identified by emerging library leaders*. <https://trends.ifla.org/update-2021>
- Jennex, Murray E. y Summer E. Bartczak. (2013). A revised knowledge pyramid. *International Journal of Knowledge Management*, 9(3), 19-30.
- Kallinikos, Jannis. (2006). Information out of information: on the self-referential dynamics of information growth. *Information technology & people*, 19(1), 98-115.
- Kao, S. C., Chang, H. C. y C. H. Lin. (2003). Decision support for the academic library acquisition budget allocation via circulation database mining. *Information Processing and Management*, 39, 133-147.



- Kim, Neung-Hoe. (2020). User experience validation using the honeycomb model in the requirements development stage. *International Journal of Advanced Smart Convergence*, 9(3), 227-231.
- Kimball, Ralph y Margy Ross. (2013). *The data warehouse toolkit: the definitive guide to dimensional modeling*. 3 ed. Indianapolis, Indiana: Wiley, Kimball Group.
- Lewis Priestley, Jennifer y Robert J. McGrath. (2019). The evolution of Data Science: a new mode of knowledge production. *International Journal of Knowledge Management*, 15(2), 1-13. doi:10.4018/IJKM.2019040106.
- Litton, Gastón, L. (1974). *La biblioteca universitaria*. Breviarios del Bibliotecario. Segunda Serie, 17. Buenos Aires, Argentina: Centro Regional de Ayuda Técnica, Agencia para el Desarrollo Internacional, Bowker.
- Lyon, Liz, Eleanor Mattern, Amelia Acker y Alison Langmead. (2015). Applying translational principles to data science curriculum development. *iPres 2015 Proceedings*. Chapel Hill, North Carolina. [http://d-scholarship.pitt.edu/27159/1/Applying\\_Translational\\_Principles\\_to\\_Dat.pdf](http://d-scholarship.pitt.edu/27159/1/Applying_Translational_Principles_to_Dat.pdf)
- Maimon, Oded y Lior Rokach. (2010). Introduction to knowledge discovery and data mining. En Oded Maimon y Lior Rokach (Eds.), *Data mining and knowledge discovery handbook*. 2 ed. (pp. 1-15) Nueva York: Springer Science+Business Media.
- Martínez-Zúñiga, Enrique y Federico Portas-Lagar. (2002). Servidor de reportes LogAleph. Guadalajara, Jal.: Biblioteca Dr. Jorge Villalobos Padilla, S. J., Instituto Tecnológico y de Estudios Superiores de Occidente. <https://rei.iteso.mx/handle/11117/659>
- Mendoza Guillén, Gabriela. (2007). *El proceso de trabajo participativo en la planeación de la evaluación de bibliotecas universitarias: estudio de caso SUBA-UACH* [Tesis de Maestría, El Colegio de México]. Repositorio COLMEX.
- Molina Félix, Luis Carlos. (1998). *Data Mining no processo de extração de conhecimento de bases de dados* [Tesis de Maestría, Universidad de São Paulo]. Biblioteca Digital USP.
- Naur, Peter. (1966). [Letters to the editor] The science of datology. *Communications of the ACM*, 9(7), 485.
- Needamangala, Ashwin. (2000). *A library decision support system built on data warehousing and data mining concepts and techniques* [Tesis de Maestría, University of Florida]. Digital Collections, George A. Smathers Libraries, University of Florida.
- Nemati, Hamid R. y Christopher D. Barko. (2003). Key factors for achieving organizational data-mining success. *Industrial Management & Data Systems*, 103(4), 282-292.
- Nicholson, Scott y Catherine Arnott Smith. (2007). Using lessons from health care to protect the privacy of library users: guidelines for the de-identification of library data based on HIPAA. *Journal of the American Society for Information Science & Technology*, 58(8), 1198-1206. <https://doi.org/10.1002/asi.20600>
- Nicholson, Scott y Jeffrey Stanton (2003). "Gaining strategic advantage through bibliomining: data mining for management decisions in corporate, special, digital, and traditional libraries."

- En Nemati, H. y C. Barko (Eds.). *Organizational data mining: Leveraging enterprise data resources for optimal performance*. Hershey, PA.: Idea Group Publishing. 247-262.
- Nicholson, Scott y Jeffrey Stanton (2006). "Bibliomining for library decision-making." En Wang, John (Ed.). *Encyclopedia of data warehousing and mining*. IGI Global. 100-105.
- Nicholson, Scott. (2000). *Creating a criterion-based information agent through data mining for automated identification of scholarly research on the world wide web* [Tesis de Doctorado, University of North Texas]. UNT Digital Library.
- Nicholson, Scott. (2003a, diciembre). The bibliomining process: data warehousing and data mining for library decision making. *Information technology and libraries*, 22(4), 146-151.
- Nicholson, Scott. (2003b). Bibliomining for automated collection development in a digital library setting using data mining to discover web-based scholarly research works. *Journal of the American Society for Information Science and Technology*. 54(12), 1081-1090.
- Nicholson, Scott. (2004). A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of documentation*, 60(2), 164-182.
- Nicholson, Scott. (2005). Digital library archaeology: a conceptual framework for understanding library use through artifact-based evaluation. *Library Quarterly*, 75(4). 496-520.
- Nicholson, Scott. (2006a). Bibliomining: Data mining for libraries: What is Bibliomining? <http://www.bibliomining.com>
- Nicholson, Scott. (2006b). Preprint of "The basis for bibliomining: frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services." *Information Processing and Management*, 42(3), 785-804.
- Nicholson, Scott. (2006c). Approaching librarianship from the data: using bibliomining for evidence-based librarianship. *Library Hi Tech*, 24(3), 369-375.
- Nicholson, Scott. (2006d, winter). Proof in the pattern: librarians follow the corporate sector toward more data-driven management. *Netconnect*, 2-6.
- OCLC. (2020). *EZproxy*. Disponible en <https://www.oclc.org/es/ezproxy.html>
- Pal, Jiban K. (2011). Usefulness and applications of data mining in extracting information from different perspectives. *Annals of Library & Information Studies*, 58(4), 7-16.
- Papatheodorou, Christos, Sarantos Kapidakis, Michalis Sfakakis y Alexandra Vassiliou. (2003). Mining user communities in digital libraries. *Information Technology and Libraries*, 22(4), 152-157.
- Pavón Plata, Armando. (2005). *Estudio de la frecuencia de uso de la colección bibliográfica general de la Biblioteca Central de la UNAM* [Tesis de Licenciatura, Universidad Nacional Autónoma de México]. TESIUNAM, Dirección General de Bibliotecas y Servicios Digitales de Información.
- Peterman, Dana, Sharon Shafer y Todd Grappone. (2021). Data lake: promoting a home-grown tool for the assessment lifecycle. *Qualitative and quantitative methods in libraries*, 10(4), 417-132.

- Portas Lagar, Federico y Álvaro Gómez Saborío. (2014). Integración de tableros de control (dash board) en la gestión bibliotecaria. En Martha Delia Castro Montoya. *Indicadores estratégicos para bibliotecas y otros servicios de información*. pp. 283-297. [http://www.sev.gob.mx/servicios/publicaciones/serie\\_fueraseries/Indicadores\\_Estrategicos\\_para\\_Bibliotecas.pdf](http://www.sev.gob.mx/servicios/publicaciones/serie_fueraseries/Indicadores_Estrategicos_para_Bibliotecas.pdf)
- Quiroz Gil, Nora Ledis y Carlos Andrés Valencia. (2012). Aplicación del proceso de KDD en el contexto de bibliomining: el caso Elogim. *Revista Interamericana de Bibliotecología*, 35(1), 97-108.
- Raina, Vineet y Srinath Krishnamurthy. (2022). *Building an effective Data Science practice: a framework to bootstrap and manage a successful Data Science practice*. Berkeley, California: Apress. [https://doi.org/10.1007/978-1-4842-7419-4\\_14](https://doi.org/10.1007/978-1-4842-7419-4_14)
- Ribas Semeler, Alexandre, Adilson Luiz Pinto, Helen Beatriz Frota Rozados. (2017). Data science in data librarianship: core competencies of a data librarian. *Journal of Librarianship and Information Science*, 51(3), 771-780.
- Rodríguez Cruz, Yunier. (2014). *Modelo de uso de información para la toma de decisiones estratégicas en organizaciones de información cubanas* [Tesis Doctoral, Universidad de Granada]. Digibug, Universidad de Granada.
- Rodríguez García, Ariel Alejandro. (2020). Retro y prospectiva de la investigación bibliotecológica sobre metadatos. En: Rodríguez García, Ariel Alejandro. *La revolución de los datos bibliográficos, científicos y culturales*. México, D. F.: Instituto de Investigaciones Bibliotecológicas y de la Información, pp. 3-16. [https://ru.ibi.unam.mx/jspui/bitstream/IIBI\\_UNAM/19/1/L232\\_Cap1.pdf](https://ru.ibi.unam.mx/jspui/bitstream/IIBI_UNAM/19/1/L232_Cap1.pdf)
- Ruiz Lobaina, Esther Marina y Pedro Lázaro Romero Suárez. (2017). Búsqueda de patrones para mejorar productos y servicios en las bibliotecas. *Investigación bibliotecológica: archivonomía, bibliotecología e información*. México, D. F.: Instituto de Investigaciones Bibliotecológicas y de la Información, 31(72), 209-225. <http://rev-ib.unam.mx/ib/index.php/ib/article/view/57830/51841>
- Sandi S., Magda Cecilia. (2019). Cincuenta años de formar profesionales en Bibliotecología en la Universidad de Costa Rica (UCR). *Revista e-Ciencias de la Información*, 9(1), 3-12.
- Secretaría de Educación Pública. (2021). *Programa para el Desarrollo Profesional Docente, para el Tipo Superior (PRODEP)*. México, D. F.: SEP. <https://dgesui.ses.sep.gob.mx/programas/programa-para-el-desarrollo-profesional-docente-para-el-tipo-superior-prodep>
- Sengupta, I. N. (1992). Bibliometrics, informetrics, scientometrics and librametrics: An overview. *Libri International Journal of Libraries Information Services*, 42(2), 75-98.
- Setién Quesada, Emilio y Salvador Gorbea Portal. (1990). Conceptos métricos en las disciplinas bibliotecario-informativas. *Actualidades de la información científica y técnica*, 21(156), 3-17.



- Showers, Ben. Editor. (2015). *Library analytics and metrics: using data to drive decisions and services*. Londres: Facet Publishing.
- Siguenza-Guzman, Lorena Catalina. (2015). *Optimal resource allocation and budgeting in libraries* [Tesis Doctoral, Ku Leuven]. Repositorio Institucional Universidad de Cuenca.
- Siguenza-Guzman, Lorena, Víctor Saquicela, Elina Ávila-Ordóñez, Joos Vandewalle y Dirk Cattrysse. (2015). "Literature review of data mining applications in academic libraries" *The Journal of Academic Librarianship*, 41, 499-510.
- Sistema Universitario de Bibliotecas Académicas. (2017). Organigrama SUBA. Chihuahua, Chih.: Universidad Autónoma de Chihuahua, Sistema Universitario de Bibliotecas Académicas. [http://suba.uach.mx/acerca\\_de/organigrama/](http://suba.uach.mx/acerca_de/organigrama/)
- Sistema Universitario de Bibliotecas Académicas. (2019). *Política de seguridad de datos*. Chihuahua, Chih. Universidad Autónoma de Chihuahua, Sistema Universitario de Bibliotecas Académicas. [https://uach.mx/assets/media/publications/2019/3/796\\_suba/Pol%C3%ADtica\\_de\\_seguridad\\_de\\_datos\\_SUBA.pdf](https://uach.mx/assets/media/publications/2019/3/796_suba/Pol%C3%ADtica_de_seguridad_de_datos_SUBA.pdf)
- Sistema Universitario de Bibliotecas Académicas. (2021a). *Directorio de bibliotecas*. Chihuahua, Chih.: Universidad Autónoma de Chihuahua, Sistema Universitario de Bibliotecas Académicas. <https://uach.mx/suba/directorio/>
- Sistema Universitario de Bibliotecas Académicas. (2021b). *Objetivos. Misión y Visión*. Chihuahua, Chih.: Universidad Autónoma de Chihuahua, Sistema Universitario de Bibliotecas Académicas. <https://uach.mx/suba/>
- Stanton, Jeffrey M. (2012). Data science: What's in it for the new librarian? *Infospace: The official blog of the Syracuse University iSchool*, 6 de julio. <https://ischool.syr.edu/infospace/2012/07/16/data-science-whats-in-it-for-the-new-librarian/>
- Su, Siew-Phek T. y Ashwin Needamangala. (2000). Harvesting information from a library data warehouse. *Information Technology in Libraries*, 19(1), 17-28.
- Tantawi, R. P. (2013). Data warehouse. *Salem Press Encyclopedia*. Research starters, EBSCOhost.
- Terrazas Guzmán, Fabiola. (04 de septiembre de 2018). Biblioteca de la DES Salud [Mensaje en un blog]. La Voz del Bibliotecario. <https://fterrazas.wordpress.com/2018/09/04/biblioteca-de-la-des-de-la-salud/>
- Universidad Autónoma de Chihuahua. (2017a). *Organigrama general UACH*. Chihuahua, Chih.: UACH. [http://transparencia.uach.mx/informacion\\_publica\\_de\\_oficio/fraccion\\_i/organigrama\\_uach\\_2017\\_1.pdf](http://transparencia.uach.mx/informacion_publica_de_oficio/fraccion_i/organigrama_uach_2017_1.pdf)
- Universidad Autónoma de Chihuahua. (2020). *Sistema de Gestión Universitario. Matriz de objetivos*. Chihuahua, Chih. UACH, 5 p.

- Universidad Autónoma de Chihuahua. (2020a). Cuarto informe anual de actividades UACH: Administración 2016-2022. Chihuahua, Chih.: UACH, 309 p. [http://transparencia.uach.mx/informacion\\_publica\\_de\\_oficio/fraccion\\_xvii/Cuarto%20informe.pdf](http://transparencia.uach.mx/informacion_publica_de_oficio/fraccion_xvii/Cuarto%20informe.pdf)
- Universidad Autónoma de Chihuahua. (2020b). *Estadística básica 2020: Administración 2016-2022*. Chihuahua, Chih.: UACH, 102 p. [https://uach.mx/assets/media/publications/2017/10/143\\_agenda-estadistica/estadistica-basica-2020-2021.pdf](https://uach.mx/assets/media/publications/2017/10/143_agenda-estadistica/estadistica-basica-2020-2021.pdf)
- Uthurusamy, Ramasamy. (1996). From data mining to knowledge discovery: current challenges and future directions. En Fayyad, Usama M., Gregory Piatetsky-Shapiro, Padhraic Smyth y Ramasamy Uthurasamy, editores. (1996). *Advances in knowledge discovery and data mining*. pp. 561-569. Menlo Park, California: American Association for Artificial Intelligence, The MIT Press.
- Villalobos Madero, Margarita y Federico Portas Lagar. (2013). *Análisis del ingreso a los recursos de información electrónica y su relación con los promedios académicos en el ITESO*. San Pedro Tlaquepaque, Jalisco: Centro de Información Académica Dr. Jorge Villalobos, S. J., ITESO.
- Wilson, Louis Round y Maurice Falcolm Tauber. (1963). *La biblioteca universitaria: su organización, administración y funciones*. Manuales del Bibliotecario, 4. Traducción: Jorge Aguayo. Washington, D. C.: Unión Panamericana.
- Xu, Dongkuan y Yingjie Tian. (2015). A comprehensive survey of clustering algorithms. *Annals of data science*, 2(2), 165-193.
- Young, Heartsill. (1988). *Glosario ALA de bibliotecología y ciencias de la información*. Traducción: Blanca de Mendizábal Allende. Madrid: Díaz de Santos.
- Zhu, Yangyong y Yun Xiong. (2015). Towards data science. *Data Science Journal*, 14(8), 1-7.

## Anexo 1. Resumen de la matriz de medición y evaluación holística de Nicholson

		TEMA	
PERSPECTIVA	<i>Sistema de biblioteca</i>	<i>Uso</i>	
<b>Interna (Sistema de biblioteca)</b>	<p><b>(1) <u>¿En qué consiste y cómo funciona el sistema de biblioteca?</u></b></p> <p>PROCESOS PROCEDIMIENTOS SERVICIOS RECURSOS COSTOS NORMAS</p> <ul style="list-style-type: none"> <li>- <i>Análisis de la estructura organizacional.</i></li> <li>- <i>Encuestas y entrevistas al personal.</i></li> <li>- <i>Evaluación de colecciones, servicios, personal, equipo/sistemas tecnológicos e instalaciones.</i></li> </ul> <p><b>Objetivo:</b> Mejorar los procedimientos y la eficiencia de la biblioteca. Determinar cuál es el desempeño del sistema desde el punto de vista de quienes laboran en él.</p>	<p><b>(4) <u>¿Cómo es utilizado el sistema de biblioteca?</u><sup>77</sup></b></p> <p>INTERACCIONES REGISTRADAS DEL USUARIO CON LOS RECURSOS Y SERVICIOS INFORMATIVOS ANÁLISIS DE LAS COLECCIONES</p> <ul style="list-style-type: none"> <li>- <i>Bibliominería.</i></li> <li>- <i>Análisis de transacciones / registros Web.</i></li> <li>- <i>Estadísticas de uso.</i></li> <li>- <i>Estadísticas de proveedores.</i></li> <li>- <i>Observación del comportamiento de los usuarios.</i></li> </ul> <p><b>Objetivo:</b> Explorar la interacción del usuario con el sistema de biblioteca para conocer sus preferencias y personalizar los servicios. Análisis de los patrones de uso.</p>	
	<b>Externa (Usuarios)</b>	<p><b>(2) <u>¿Qué tan efectivo es el sistema de biblioteca?</u></b></p> <p>CALIDAD DE COLECCIONES Y SERVICIOS RELEVANCIA PERTINENCIA ESTUDIOS DE USABILIDAD</p> <ul style="list-style-type: none"> <li>- <i>Encuestas y entrevistas a usuarios (sobre colecciones físicas y digitales, configuración de instalaciones, organización de recursos físicos, señalética, facilidad de uso de las interfaces de usuario, evaluación de UI y UX etc.).</i></li> <li>- <i>Mecanismos de comunicación en voz alta y retroalimentación durante el proceso.</i></li> <li>- <i>Recopilación de estadísticas.</i></li> <li>- <i>Buzón de sugerencias.</i></li> <li>- <i>Grupos focales.</i></li> </ul> <p><b>Objetivo:</b> Dimensionar la calidad y el desempeño del sistema desde el punto de vista o percepción del usuario (satisfacción de sus necesidades de información).</p>	<p><b>(3) <u>¿Qué tan útil resulta el sistema de biblioteca al usuario?</u></b></p> <p>ESTADOS DEL CONOCIMIENTO SATISFACCIÓN DE NECESIDADES INFORMATIVAS EXPERIENCIA DE USO CITAS DEL USUARIO A RECURSOS INFORMATIVOS</p> <ul style="list-style-type: none"> <li>- <i>Encuestas y entrevistas a usuarios.</i></li> <li>- <i>Grupos focales.</i></li> <li>- <i>Análisis de citas de los materiales hechas por los usuarios (Aplicación de los EMI).</i></li> </ul> <p><b>Objetivo:</b> Determinar cuál es el punto de vista del usuario sobre la utilidad del sistema a partir del uso que hace del mismo. Analizar el impacto del sistema de biblioteca en los usuarios.</p>

Fuente: Elaboración propia fundamentada en Nicholson (2014, p.167-174) y Siguenza-Guzman (2015). (Traducción propia).

<sup>77</sup> De acuerdo con Nicholson, los estudios referidos a este cuadrante “No descubren lo que el usuario quería hacer, ni lo que podría haber hecho, sin embargo, captura lo que el usuario realmente hizo (en comparación a lo que el usuario dice haber hecho)”. Nicholson, 2004, p. 172. (Traducción propia).

## Anexo 2. Proveedores de servicios de información especializada para la toma de decisiones en bibliotecas

País	Empresa	Servicio	Página electrónica
Colombia	Elogim SAS	Elogim	<a href="https://elogim.com/">https://elogim.com/</a>
Estados Unidos	ProQuest - Ex Libris	PRIMO Analytics	<a href="https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/Primo/Analytics">https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/Primo/Analytics</a>
Estados Unidos	ProQuest - Ex Libris	ALMA Analytics	<a href="https://exlibrisgroup.com/es/productos/alma/analytics/">https://exlibrisgroup.com/es/productos/alma/analytics/</a>
Estados Unidos	OCLC	EZProxy Analytics	<a href="https://www.oclc.org/es/ezproxy/ezproxy-analytics.html">https://www.oclc.org/es/ezproxy/ezproxy-analytics.html</a>
Estados Unidos	EBSCO	Panorama	<a href="https://www.ebsco.com/products/panorama">https://www.ebsco.com/products/panorama</a>
Estados Unidos	RedLink> An Atypon Company	Library Dashboard by RedLink>	<a href="https://redlink.com/library-dashboard/">https://redlink.com/library-dashboard/</a>
México	ITMS Analytics Inc.	ITMS Analytics	<a href="https://www.itmsanalytics.com/">https://www.itmsanalytics.com/</a>
México	Infoestratégica	InfoIntelligen	<a href="http://www.infointelligen.com/">http://www.infointelligen.com/</a>

Fuente: Elaboración propia.

### **Anexo 3. Normas de operación del SUBA-UACH**

Para cumplir con su misión y visión, el SUBA ha establecido las siguientes normas de operación:

1. Satisfacer las necesidades de información de los usuarios, en cumplimiento con los fines de la Universidad y basados en los principios de libre acceso, legalidad, eficacia, eficiencia, transparencia y calidad en la prestación de servicios.
2. Ofrecer programas de capacitación y actualización al personal del SUBA, de forma sistemática y continua, compartiendo los resultados de sus investigaciones y experiencias.
3. Trabajar en forma integrada con los miembros de la comunidad universitaria, en el cumplimiento de los fines esenciales de la Universidad.
4. Evaluar de manera continua la colección documental para detectar fortalezas y debilidades, con el fin de contar con elementos para realizar una selección adecuada.
5. Desarrollar y mantener una colección adecuada, suficiente y actualizada, que cubra todos los programas académicos de nuestra institución.
6. Administrar y custodiar el Archivo Histórico Universitario.
7. Administrar y custodiar las colecciones especiales de las Bibliotecas Académicas.
8. Administrar y custodiar la colección especial de todo el material documental que se genere en nuestra Universidad.
9. Motivar y estimular el hábito de la lectura y la investigación, así como fomentar en la comunidad, cultura para el buen uso y aprovechamiento de los recursos bibliotecarios y proporcionar al usuario un ambiente adecuado (Sistema Universitario de Bibliotecas Académicas, 2016).

Fuente: Página electrónica del SUBA-UACH.

## Anexo 4. Encuesta sobre Conducta Informativa de Profesionales de la Información

¡Buen día! Gracias por participar en la Encuesta CI-PI, que tiene como objetivo el conocer la conducta informativa de profesionales de la información en la toma de decisiones y solución de problemas. Esta encuesta forma parte del proyecto de tesis "Diseño de un modelo de bibliominería para identificar patrones de uso que generen conocimiento significativo para la toma de decisiones y la evaluación de actividades bibliotecarias". Incluye preguntas sobre las acciones que los profesionales de la información llevan a cabo para obtener información que necesitan para la toma de decisiones y solución de problemas que se presentan en su trabajo habitual en centros de información, específicamente en bibliotecas. Le tomará aproximadamente 10 minutos completar el cuestionario. Agradeceré mucho que aporte su experiencia al respecto para enriquecer esta investigación. Las respuestas obtenidas serán estrictamente confidenciales. Si tiene dudas con respecto a la encuesta puede comunicarse con Gabriela Mendoza, al correo electrónico [gabriela.mendoza@outlook.com](mailto:gabriela.mendoza@outlook.com). Muchas gracias por su tiempo y apoyo al proyecto.

1. Solicitud de información de contacto
2. ¿A qué grupo de edad pertenece?
  - 20-24 años
  - 25-29 años
  - 30-34 años
  - 35-39 años
  - 40-44 años
  - 46-49 años
  - 50-54 años
  - 55-59 años
  - 60-64 años
  - 65-69 años
  - 70-74 años
  - 75-79 años
3. Sexo
  - Femenino
  - Masculino
4. Indique su máximo grado de estudios (aprobado):
  - Primaria
  - Secundaria
  - Bachillerato
  - Licenciatura
  - Maestría
  - Doctorado
  - Posdoctorado
  - Especialidad
5. ¿Qué idiomas domina? (Seleccione los que apliquen)
  - Español
  - Inglés

- Francés
- Portugués
- Otro: \_\_\_\_\_

6. Indique el nombre de su centro de trabajo: \_\_\_\_\_

7. ¿Cuál es su puesto de trabajo actual? \_\_\_\_\_

8. Por favor describa brevemente el tipo de trabajo que realiza, enfatizando cuáles son sus principales tareas y actividades: \_\_\_\_\_

9. ¿Cuenta con estudios en Bibliotecología y Documentación?

- Sí ¿Cuál(es)? \_\_\_\_\_
- No

10. ¿Cuánto tiempo de experiencia tiene trabajando en biblioteca?

- 1-3 años
- 4-6 años
- 7-9 años
- 10-12 años
- 13-15 años
- 16-18 años
- 19-21 años
- 22-24 años
- 25-27 años
- 28 años o más

11. ¿Lee regularmente alguna revista o publicación especializada relacionada con su trabajo?

- Sí ¿Cuál(es)? \_\_\_\_\_
- No

12. Por favor mencione hasta tres temas que considere son los más importantes con respecto a su trabajo en biblioteca:

- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_

13. ¿Dónde consigue la mayor parte de la información que requiere sobre estos temas?

\_\_\_\_\_

14. Por favor mencione todas las fuentes de información que consulta habitualmente para resolver problemas y/o tomar decisiones en su trabajo diario:

\_\_\_\_\_

15. ¿Pertenece a sociedades, asociaciones profesionales u organizaciones relacionadas con su trabajo?

- Sí ¿Cuál(es)? \_\_\_\_\_
- No

16. Dónde obtiene información útil para resolver problemas y/o tomar decisiones en su área de trabajo? Por favor marque todas las opciones válidas:

- Manuales de procedimiento internos
- Asistiendo a reuniones de trabajo
- Minutas de reuniones de trabajo
- Consultando a colegas
- Investigando en literatura especializada en Bibliotecología
- Sistema de gestión de biblioteca – *Library Management System (LMS) – Integrated Library System (ILS)*
- Manuales y reglas de catalogación y clasificación
- Buscando en Internet
- Cursos y/o talleres de capacitación
- Redes sociales
- Listas de discusión en correo electrónico
- Participando en encuentros profesionales del gremio bibliotecológico
- Legislación o normativa en la materia (Leyes, reglamentos, políticas)
- Planes, programas y proyectos institucionales
- Informes de labores de administraciones pasadas
- Experiencia de trabajo propia en otros centros de información
- Consulta a usuarios de biblioteca (entrevistas, encuestas, etc.)
- Alertas de información (RSS)
- Bases de datos especializadas
- Libros de texto especializados
- Publicaciones periódicas especializadas
- Otra(s) fuente(s): \_\_\_\_\_

17. ¿Cuál es su nivel de FAMILIARIDAD con cada una de las siguientes fuentes de información:

Fuente	No familiarizado	No muy familiarizado	Algo familiarizado	Muy familiarizado	Extremadamente familiarizado
Manuales de procedimiento internos					
Reuniones de trabajo					
Minutas de reuniones de trabajo					
Estadísticas del trabajo en bibliotecas					
Literatura especializada en bibliotecología (impresa)					
Literatura especializada en bibliotecología (digital)					
Sistema de gestión de biblioteca (LMS/ILS)					
Manuales y reglas de catalogación y clasificación					
Cursos y/o talleres de capacitación					
Búsqueda en Internet					
Redes sociales					
Listas de discusión en correo electrónico					
Encuentros profesionales del gremio bibliotecológico					
Legislación o normativa en materia bibliotecológica					
Planes, programas y proyectos institucionales					
Consulta a usuarios de bibliotecas (entrevistas, encuestas, etc.)					
Consulta a colegas					
Alertas de información (RSS)					
Bases de datos especializadas					



18. Considerando todo lo que ha escuchado/conocido sobre las fuentes de información, independientemente de que las haya empleado o no ¿qué tan RELEVANTES le resultan cada una de las siguientes fuentes (si no conoce alguna de ellas, por favor seleccione NA)?

Fuente	No relevante	No muy relevante	Algo relevante	Muy relevante	Extremadamente relevante
Manuales de procedimiento internos					
Reuniones de trabajo					
Minutas de reuniones de trabajo					
Estadísticas del trabajo en bibliotecas					
Literatura especializada en bibliotecología (impresa)					
Literatura especializada en bibliotecología (digital)					
Sistema de gestión de biblioteca (LMS/ILS)					
Manuales y reglas de catalogación y clasificación					
Cursos y/o talleres de capacitación					
Búsqueda en Internet					
Redes sociales					
Listas de discusión en correo electrónico					
Encuentros profesionales del gremio bibliotecológico					
Legislación o normativa en materia bibliotecológica					
Planes, programas y proyectos institucionales					
Consulta a usuarios de bibliotecas (entrevistas, encuestas, etc.)					
Consulta a colegas					
Alertas de información (RSS)					
Bases de datos especializadas					

19. Considerando todo lo que ha escuchado/conocido sobre las fuentes de información, independientemente de que las haya empleado o no ¿qué tan CONFIABLES le resultan cada una de las siguientes fuentes (si no conoce alguna de ellas, por favor seleccione NA)?

Fuente	No confiable	No muy confiable	Algo confiable	Muy confiable	Extremadamente confiable
Manuales de procedimiento internos					
Reuniones de trabajo					
Minutas de reuniones de trabajo					
Estadísticas del trabajo en bibliotecas					
Literatura especializada en bibliotecología (impresa)					
Literatura especializada en bibliotecología (digital)					
Sistema de gestión de biblioteca (LMS/ILS)					
Manuales y reglas de catalogación y clasificación					
Cursos y/o talleres de capacitación					
Búsqueda en Internet					
Redes sociales					
Listas de discusión en correo electrónico					
Encuentros profesionales del gremio bibliotecológico					
Legislación o normativa en materia bibliotecológica					
Planes, programas y proyectos institucionales					
Consulta a usuarios de bibliotecas (entrevistas, encuestas, etc.)					
Consulta a colegas					
Alertas de información (RSS)					
Bases de datos especializadas					

20. Considerando todo lo que ha escuchado/conocido sobre las fuentes de información, independientemente de que las haya empleado o no ¿qué tan VALIOSAS considera que le resultan (o pudieran resultar) tales fuentes en el desempeño de su trabajo diario (si no conoce alguna de ellas, por favor seleccione NA)?

Fuente	No valiosa	No muy valiosa	Algo valiosa	Muy valiosa	Extremadamente valiosa
Manuales de procedimiento internos					
Reuniones de trabajo					
Minutas de reuniones de trabajo					
Estadísticas del trabajo en bibliotecas					
Literatura especializada en bibliotecología (impresa)					
Literatura especializada en bibliotecología (digital)					
Sistema de gestión de biblioteca (LMS/ILS)					
Manuales y reglas de catalogación y clasificación					
Cursos y/o talleres de capacitación					
Búsqueda en Internet					
Redes sociales					
Listas de discusión en correo electrónico					
Encuentros profesionales del gremio bibliotecológico					
Legislación o normativa en materia bibliotecológica					
Planes, programas y proyectos institucionales					
Consulta a usuarios de bibliotecas (entrevistas, encuestas, etc.)					
Consulta a colegas					
Alertas de información (RSS)					
Bases de datos especializadas					

21. Considerando todo lo que ha escuchado/conocido sobre las fuentes de información, independientemente de que las haya empleado o no ¿qué tanto INTERÉS posee en emplearlas en su trabajo diario (si no conoce alguna de ellas, por favor seleccione NA)?

Fuente	Sin interés	No mucho interés	Algo de interés	Con interés	Mucho interés
Manuales de procedimiento internos					
Reuniones de trabajo					
Minutas de reuniones de trabajo					
Estadísticas del trabajo en bibliotecas					
Literatura especializada en bibliotecología (impresa)					
Literatura especializada en bibliotecología (digital)					
Sistema de gestión de biblioteca (LMS/ILS)					
Manuales y reglas de catalogación y clasificación					
Cursos y/o talleres de capacitación					
Búsqueda en Internet					
Redes sociales					
Listas de discusión en correo electrónico					
Encuentros profesionales del gremio bibliotecológico					
Legislación o normativa en materia bibliotecológica					
Planes, programas y proyectos institucionales					
Consulta a usuarios de bibliotecas (entrevistas, encuestas, etc.)					
Consulta a colegas					
Alertas de información (RSS)					
Bases de datos especializadas					

22. ¿Cómo calificaría la ACCESIBILIDAD/DISPONIBILIDAD de las siguientes fuentes de información en su área de trabajo diario?

Fuente	No accesible/disponible	Rara vez accesible/disponible	Habitualmente accesible/disponible	Siempre accesible/disponible
Manuales de procedimiento internos				

Fuente	No accesible/disponible	Rara vez accesible/disponible	Habitualmente accesible/disponible	Siempre accesible/disponible
Reuniones de trabajo				
Minutas de reuniones de trabajo				
Estadísticas del trabajo en bibliotecas				
Literatura especializada en bibliotecología (impresa)				
Literatura especializada en bibliotecología (digital)				
Sistema de gestión de biblioteca (LMS/ILS)				
Manuales y reglas de catalogación y clasificación				
Cursos y/o talleres de capacitación				
Búsqueda en Internet				
Redes sociales				
Listas de discusión en correo electrónico				
Encuentros profesionales del gremio bibliotecológico				
Legislación o normativa en materia bibliotecológica				
Planes, programas y proyectos institucionales				
Consulta a usuarios de bibliotecas (entrevistas, encuestas, etc.)				
Consulta a colegas				
Alertas de información (RSS)				
Bases de datos especializadas				

23. Desde su punto de vista ¿En qué aspectos radica la UTILIDAD de un recurso de información? Por favor enumere del uno al trece en orden de importancia (donde uno es lo más importante y trece lo menos importante):

- \_\_\_ Vigencia
- \_\_\_ Confiabilidad (prestigio del autor/editorial/publicación)
- \_\_\_ Representatividad en área temática
- \_\_\_ Disponibilidad
- \_\_\_ Temas/Contenido
- \_\_\_ Actualidad
- \_\_\_ Pertinencia
- \_\_\_ Formato
- \_\_\_ Originalidad
- \_\_\_ Alcance/Audiencia
- \_\_\_ Número de préstamos
- \_\_\_ Costo
- \_\_\_ Utilidad

Fuente: Elaboración propia.

## Anexo 5. Parámetros y variables incluidos en cada categoría de reportes del LogAleph del SUBA-UACH

### CATEGORÍA USO DEL ACERVO

USO DEL ACERVO		USO DEL ACERVO	
		Variables	Frecuencia
		Total	31
		#	49
		Tipo de usuario	18
		Transacción	18
		# Invent	17
		Biblioteca	15
		Bibliotecario	13
		Clasificación	13
		Título	13
		Autor	11
		Colección	11
		Editorial	11
		Matrícula	11
		Fecha	10
		Hora	10
		Usuario	10
		Tipo de acervo	9
		Tipo de préstamo	9
		Descrip	8
		Año (AÁ'O)	6
		Material	6
		Sexo Hombre(H)/Mujer(M) - Alumnos UACH-DS	3
		Carrera	2
		Clave de la carrera - Divison (Alumnos UACH-DS)	2
		Estatus del MD	2
		Fecha de actualización	2
		Número de préstamos	2
		Número de títulos	2
		Sexo Hombre(H)/Mujer(M) - Formacion (Alumnos UACH-DS)	2
		Última fecha de préstamo	2
		Código de barras	1
		Código de barras (Matrícula)	1
		Fecha de creación	1
		Fecha de creación (FECHA DE CREACIÃ"N)	1
		Fecha de vencimiento	1
		Fecha en que se reservó	1
		Fecha préstamo	1
		Fecha vencimiento	1
		Hora en que se reservó	1
		ID Usuario	1
		No. De Usuarios	1
		Número de matrícula	1
		Número de sistema	1
		OBSERVACIONES,(1),(2),(3),(4),(5)'	1
		P_ Externo	1
		P_ Interno	1
		Sexo Hombre(H)/Mujer(M) Alumnos UACH-DS	1

USO DEL ACERVO	
Parámetros	Frecuencia
Biblioteca	40
Fecha final (año/mes/día)	31
Fecha inicial (año/mes/día)	31
DIRECTO (todas las bibliotecas)	4
Hora de inicio	2
Hora de término	2
Matrícula	2
Status	2
Desde (año/mes/día)	1
Hasta (año/mes/día)	1
Nombre de la cuenta staff bibliotecario	1
Número de IP	1

## **CATEGORÍA MATERIAL VENCIDO**

<b>MATERIAL VENCIDO</b>		<b>MATERIAL VENCIDO</b>	
<b>Parámetros</b>	<b>Frecuencia</b>	<b>Variables</b>	<b>Frecuencia</b>
Biblioteca	5	#	6
Fecha final (año/mes/día)	4	Cantidad	6
Fecha inicial (año/mes/día)	4	Matrícula	6
Fecha (año/mes/día)	1	Transacción	5
Matrícula	1	Usuario	5
		Bibliotecario	4
		Descrip	4
		Carrera	3
		Crédito o débito	3
		Fecha y hora	3
		# Invent	1
		Biblioteca	1
		Fecha	1
		Fecha préstamo	1
		Fecha vencimiento	1
		ID Usuario	1

## **CATEGORÍA SERVICIOS BIBLIOTECARIOS**

<b>SERVICIOS BIBLIOTECARIOS</b>		<b>SERVICIOS BIBLIOTECARIOS</b>	
<b>Parámetros</b>	<b>Frecuencia</b>	<b>Parámetros</b>	<b>Frecuencia</b>
Fecha final (año/mes/día)	3	#	5
Fecha inicial (año/mes/día)	3	Total	4
Biblioteca	2	Biblioteca	3
		Transacción	3
		# Invent	1
		Año	1
		Clasificación	1
		Colección	1
		Editorial	1
		Número de préstamos	1
		Sexo Hombre(H)/Mujer(M)-FORMACION (Alumnos UACH-DS)	1
		Título	1
		Última fecha de préstamo	1

**CATEGORÍA UNIDAD DE SISTEMAS**

<b>UNIDAD DE SISTEMAS</b>	
<b>Parámetros</b>	<b>Frecuencia</b>
Biblioteca	4
Fecha final (año/mes/día)	2
Fecha inicial (año/mes/día)	2

<b>UNIDAD DE SISTEMAS</b>	
<b>Variables</b>	<b>Frecuencia</b>
Usuario	5
#	4
Bibliotecario	4
Código de barras (Matrícula)	3
Carrera	2
Clave de la carrera - Division (alumnos UACH-DS)	2
Descrip	2
Fecha creación	2
Fecha de actualización	2
Fecha de creación	2
Fecha pago	2
Folio	2
Importe	2
IP	2
Nombre	2
Sexo Hombre(H)/Mujer(M) - Formacion (Alumnos UACH-DS)	2
Transacción	2
Biblioteca	1
Sexo Hombre(H)/Mujer(M) - Alumnos UACH-DS	1
Z303_Note_1	1
Z303_First_Name	1
Z303_Last_Name	1
Z303_Note_2	1
Z303_User_Library	1
Z304_Email_Address	1
Z305_Field_1	1
Z305_Field_2	1
Z305_Field_3	1
Z305_No_Photo	1
Z305_Update_Date	1

## CATEGORÍA PROCESOS TÉCNICOS

PROCESOS TÉCNICOS	
Parámetros	Frecuencia
Biblioteca	34
Desde (año/mes/día)	27
Hasta (año/mes/día)	27
Catalogador	10
Fecha final (año/mes/día)	7
Fecha inicial (año/mes/día)	7
NA	5
Texto	3
Dewey final	2
Dewey inicial	2
Ejemplar	2
ERROR	1
Estatus	1
Hora de inicio:	1
Hora de término:	1
Matrícula	1

PROCESOS TÉCNICOS	
Variables	Frecuencia
#	55
Catalogador	23
Clasificación	21
Título	21
# Invent	20
Descrip	20
Autor	19
Colección	18
Estatus del MD	15
Material	14
Z30 Item Statistic	13
Número de sistema	10
Razón	9
Biblioteca	8
Total	8
Fecha	7
Ítem	6
Tipo de préstamo	5
Carrera	4
Fecha apertura	4
Matrícula	4
Usuario	4
Código de barras (matrícula)	3
Editorial	3
Fecha de actualización	3
Fecha de creación	3
Fecha de vencimiento	3
Número de sistema - Ítems	3
Sexo Hombre (H)/Mujer (M) - Formación (Alumnos UACH-DS)	3
Tipo de usuario	3
Bibliotecario	2
Cantidad	2
Clave de la carrera - División (Alumnos UACH-DS)	2
Fecha actualización	2
Número de ítem	2
Z30H_H_Hour	2
C.B-Escaneado-Aquí'	1
Clave de la carrera {	1
Crédito o débito	1
Dirección IP	1
Ejemplares	1
Fecha y hora	1
Modificaciones o bajas	1
Observaciones, (1),(2),(3),(4),(5)'	1
Substr(Z13_Imprint,-5,5)	1
Substr(Z308_Rec_Key,3,9)	1
Tipo de acervo	1
Títulos	1
Z103_Chronological_I	1
Z103_Chronological_J	1
Z103_Enumeration_A	1
Z103_Enumeration_B	1
Z103_Enumeration_C	1
Z103_LKR_Library	1
Z103_LKR_Text_M	1
Z103_LKR_Text_M_Alpha	1
Z103_LKR_Text_N	1
Z103_LKR_Text_N_Alpha	1
Z103_LKR_Type	1
Z103_Rec_Key	1
Z103_Rec_Key_1	1
Z103_Sort	1
Z31_Key_Type	1
Z31_Type	1

Fuente: Elaboración propia a partir de la revisión de los reportes vigentes del LogAleph.

## Anexo 6. Variables de interés vs variables disponibles en la BDESSalud

Para la delimitación y categorización de las variables de interés que permiten llevar a cabo labores de bibliominería se consideró útil ver cuáles de ellas daban respuesta a las preguntas de qué, quién, cómo, cuándo, dónde y porqué<sup>78</sup>:

### ¿Qué recurso han sido utilizado?

Aquí se hace referencia al MDoc de la biblioteca bajo estudio, las variables dan la posibilidad de llevar a cabo labores de identificación y clasificación y/o agrupación de materiales (tanto físicos como electrónicos) de acuerdo con sus patrones transaccionales (circulación): títulos más consultados, materias o temas prioritarios, tipo de materiales preferidos, material con cero uso, material con mayor número de préstamos y/o renovaciones y/o vencimientos, tipo de adquisición, idiomas preferidos, etc.

VARIABLES DESEABLES Recurso utilizado	VARIABLES DISPONIBLES BDESSalud
ID MDoc	Sí
Clasificación	Sí
ISBN	Sí
ISSN	Sí
Autor	Sí
Título	Sí
Año de publicación	Sí
Edición	No
Editorial	Manejan pie de imprenta
Idioma	No
Tema / Materia	Fue obtenida a partir de la clasificación
Tipo de material	Sí
Tipo de colección	Sí
Tipo de adquisición	Sí
Proveedor	No
Fecha de adquisición	No
Precio MDoc	No
Tipo de donación	No
Tipo de donante	No
Estatus del MDoc	Sí
Cubículos	Desarrollo incipiente
Cursos	Desarrollo incipiente
Equipo de cómputo	Desarrollo incipiente
Instalaciones físicas de la biblioteca	No
Atriles	Desarrollo incipiente
Página web o dominio URL de la biblioteca	No
Catálogo en línea	No

<sup>78</sup> Aplicadas a los elementos involucrados en las transacciones y servicios bibliotecarios de tipo académico.



<b>VARIABLES DESEABLES</b> <b>Recurso utilizado</b>	<b>VARIABLES DISPONIBLES BDESSalud</b>
Descubridor de recursos	No
Bases de datos	No
Artículo de Base de Datos (Texto completo)	No
Artículo de Base de Datos (Referencia)	No
Libro de Base de Datos (Texto completo)	No
Libro de Base de Datos (Referencia)	No
Recurso digital	No
RRSS de la biblioteca	No
Repositorios digitales de la institución	No

### **¿Qué transacción se ha realizado en el sistema?**

<b>VARIABLES DESEABLES</b> <b>Transacción realizada</b>	<b>VARIABLES DISPONIBLES BDESSalud</b>
ID Transacción	Sí
Préstamo externo	Sí
Préstamo interno (uso en sala)	Sí
Renovación de MDoc hecha por bibliotecario	Sí
Renovación de MDoc hecha por usuario	Sí
Devolución de MDoc	Sí
Apartado de MDoc	Desarrollo incipiente
Préstamo interbibliotecario	No
Préstamo vigente	No
Préstamo vencido	Sí
Multa por devolución tardía	Sí
Pago de multa (Usuario)	Sí
Cobro de multa (Bibliotecario)	Sí
Cancelación de multa (Bibliotecario)	Sí
Ingreso a sistema (Registro de bibliotecario)	No
Modificación al registro de bibliotecario	No
Baja del sistema (Bibliotecario)	No
Solicitud de compra de MDoc	No
Ingreso a sistema (Registro de proveedor)	No
Modificación al registro de proveedor	No
Baja del sistema (Proveedor)	No
Adquisición de MDoc	No
Ingreso a sistema (Registro de MDoc)	Sí
Proceso técnico del MDoc	No
Modificación al registro de MDoc	Sí
Baja de sistema (MDoc)	Sí
Ingreso a sistema (Registro de usuario)	Sí
Modificación al registro de usuario	Sí
Baja del sistema (Usuario)	Sí
Préstamo de ítem distinto al MDoc	No

<b>VARIABLES DESEABLES</b> <b>Transacción realizada</b>	<b>VARIABLES DISPONIBLES BDESSalud</b>
Préstamo de cubículo	Desarrollo incipiente
Asistencia a curso/taller/evento/actividad de biblioteca	Desarrollo incipiente
Servicio de referencia	No
Ingreso al dominio o página web de la biblioteca	No
Interacción/búsqueda en el catálogo de biblioteca	No
Acceso e interacción con el servicio de descubrimiento de recursos	No
Acceso e interacción con la plataforma de servicios bibliotecarios	No
Acceso e interacción con las bases de datos	No
Acceso e interacción con repositorios digitales	No
Descarga de artículo en texto completo	No
Consulta con resultados	No
Consulta fallida (sin resultados)	No
Términos empleados en la búsqueda	No
Número de consultas electrónicas	No
Interacción con RRSS de biblioteca	No

**¿Quién ha realizado alguna transacción en el sistema de la biblioteca?  
(Préstamo/Renovación/Devolución/Apartado/Multas/Acceso/Descarga/etc.)**

<b>VARIABLES GENÉRICAS DESEABLES</b> <b>Usuario</b>	<b>VARIABLES DISPONIBLES BDESSalud</b>
ID Usuario	Sí
Fecha de nacimiento	No
Sexo	Sí
Tipo/Categoría de usuario	Sí
Vigencia/Estatus	Sí
Fecha de creación de registro de usuario	Sí
Fecha de actualización de registro de usuario	Sí
Fecha de eliminación de registro de usuario	Sí
Biblioteca de adscripción	Sí
Campus	No
Unidad académica/Facultad/Dependencia académica	Se obtuvo empleando la clave de carrera
Departamento	No
Carrera/Plan de estudios	Sí

**Variables especializadas por tipo de usuario:**

<b>VARIABLES DESEABLES PARA Estudiante</b>	<b>VARIABLES DISPONIBLES BDESSalud</b>
ID Usuario Estudiante	Sí
Fecha de nacimiento	No
Sexo	Sí
Tipo/Categoría de usuario	Sí
Vigencia/Estatus	Sí
Fecha de creación de registro de usuario	Sí
Fecha de actualización de registro de usuario	Sí
Fecha de eliminación de registro de usuario	Sí
Biblioteca de adscripción	Sí
Campus	No
Unidad académica/Facultad/Dependencia académica	Se obtuvo empleando la clave de carrera
Departamento	No
Carrera/Plan de estudios	No
Nivel o grado académico en curso	No
Promedio académico (calificaciones)	No
Nivel de inglés	No
¿Recibe alguna beca?	No
Semestre	No
Semestres cursados	No
Créditos requeridos	No
Créditos aprobados	No
Créditos inscritos	No
Créditos por cursar	No
Ingreso al plan de estudios	No
Último periodo cursado	No
Pertenencia al PNPC	No
Egreso	No
Puntaje Examen General de Egreso de Licenciatura (EGEL)	No
Titulación	No

<b>VARIABLES DESEABLES PARA Docente/Académico/Catedrático/Investigador</b>	<b>VARIABLES DISPONIBLES BDESSalud</b>
ID Usuario Docente	Sí
Fecha de nacimiento	No
Sexo	Sí
Tipo/Categoría de usuario	Sí
Vigencia/Estatus	Sí
Fecha de creación de registro de usuario	Sí
Fecha de actualización de registro de usuario	Sí
Fecha de eliminación de registro de usuario	Sí
Biblioteca de adscripción	Sí

<b>VARIABLES DESEABLES PARA Docente/Académico/Catedrático/Investigador</b>	<b>VARIABLES DISPONIBLES BDESSalud</b>
Campus	No
Unidad académica/Facultad/Dependencia académica	Se obtuvo empleando la clave de carrera
Departamento	No
Carrera/Plan de estudios en que labora	No
Nivel o grado académico máximo obtenido	No
Nivel de inglés	No
Categoría (Tiempo Completo, Medio Tiempo, Hora Clase, etc.)	No
Antigüedad	No
Recibe capacitación institucional	No
Resultados en la evaluación de la práctica docente	No
Pertenencia a algún Cuerpo Académico de Investigación	No
Perfil del Sistema Nacional de Investigación (SIN)	No
Perfil deseable PRODEP	No
Perfil deseable PROMEP	No
Recibe otros estímulos	No
Currículum Vitae <u>con publicaciones incluidas</u>	No

<b>VARIABLES DESEABLES PARA Personal administrativo</b>	<b>VARIABLES DISPONIBLES BDESSalud</b>
ID Usuario Administrativo	Sí
Fecha de nacimiento	No
Sexo	Sí
Tipo/Categoría de usuario	Sí
Vigencia/Estatus	Sí
Fecha de creación de registro de usuario	Sí
Fecha de actualización de registro de usuario	Sí
Fecha de eliminación de registro de usuario	Sí
Biblioteca de adscripción	No
Campus	No
Unidad académica/Facultad/Dependencia académica	Se obtuvo empleando la clave de carrera
Departamento	No
Nivel o grado académico máximo obtenido	No
Nivel de inglés	No
Categoría (Sindicalizado/De confianza)	No
Antigüedad	No

### ¿Quién atendió la transacción/ofreció el servicio?

Personal bibliotecario que realizó la transacción

VARIABLES DESEABLES PARA Personal bibliotecario	VARIABLES DISPONIBLES BDESSalud
ID Bibliotecario	Sí
Fecha de nacimiento	No
Sexo	No
Categoría (Sindicalizado/De confianza)	No
Puesto de trabajo	No
Vigencia/Estatus	No
Fecha de creación de registro de bibliotecario	No
Fecha de actualización de registro de usuario	No
Fecha de eliminación de registro de bibliotecario	No
Biblioteca de adscripción	No
Campus	No
Unidad académica/Facultad/Dependencia académica	No
Departamento	No
Nivel o grado académico máximo obtenido	No
Antigüedad	No
Ha recibido capacitación en sus labores bibliotecarias	No

### ¿Cómo se efectuó la transacción?

¿Qué características tienen esas transacciones?

VARIABLES DESEABLES Características de la transacción	VARIABLES DISPONIBLES BDESSalud
Devolución temprana de MDoc	Es factible calcularla
Devolución en tiempo de MDoc	Es factible calcularla
Devolución tardía de MDoc	Es factible calcularla
Errores de devolución de MDoc	Sí
Errores en proceso técnico de MDoc	No
Análisis de proxys	No
Número de consultas electrónicas	No
Acciones emprendidas por usuario final	No
Búsqueda con resultados	No
Términos de búsqueda más populares	No
Descarga de artículo en texto completo	No
Búsqueda fallida (con 0 resultados)	No
Términos empleados en búsqueda fallida	No
Duración de la sesión de uso de recursos electrónicos	No
Tipo de dispositivo empleado en la sesión	No
Tipo de interacción con RRSS de biblioteca (positiva/negativa)	No
Número de personas que usaron cubículos	No

<b>VARIABLES DESEABLES</b> <b>Características de la transacción</b>	<b>VARIABLES DISPONIBLES BDESSalud</b>
Grado de satisfacción de los usuarios con los servicios bibliotecarios	No
Criterios de encuestas de satisfacción con los servicios	No
Resultados de evaluación a curso/taller de biblioteca	No
Resultados de evaluación a evento de biblioteca	No

### **¿Cuándo se efectuó la transacción?**

Temporalidad de la transacción

<b>VARIABLES DESEABLES</b>	<b>VARIABLES DISPONIBLES BDESSalud</b>
Fecha transacción (AAAAMMDD)	Sí
Fecha transacción (dato textual)	No
Día transacción (dato numérico)	Sí
Día transacción (dato textual)	No
Mes transacción (dato numérico)	No
Mes transacción (dato textual)	No
Año transacción (dato numérico)	No
Año transacción (dato textual)	No
Hora transacción (HHMM)	Sí
Semestre	No
Trimestre	No
Cuatrimestre	No
Bimestre	No
Semana	No
Día hábil	No
Fecha de ingreso a sistema (MDoc)	Sí
Fecha de ingreso a sistema (Usuario)	Sí
Fecha de ingreso a sistema (Bibliotecario)	No
Fecha de ingreso a sistema (Proveedor)	No
Fecha de actualización (MDoc)	Sí
Fecha de actualización (Usuario)	Sí
Fecha de actualización (Bibliotecario)	No
Fecha de actualización (Proveedor)	No
Fecha de baja (MDoc)	Sí
Fecha de vencimiento de registro (Usuario)	Sí
Fecha de vencimiento de registro (Bibliotecario)	No
Fecha de vencimiento de registro (Proveedor)	No
Fecha de la última transacción (MDoc)	Sí
Fecha de la última transacción (Usuario)	No
Fecha de la última transacción (Bibliotecario)	No
Fecha de la última transacción (Proveedor)	No
Última fecha de préstamo (MDoc)	Sí

### ¿Dónde se efectuó la transacción?

<b>VARIABLES DESEABLES</b> <b>Ubicación de la transacción</b>	<b>VARIABLES DISPONIBLES BDESSalud</b>
Registro de ingreso a las instalaciones de la biblioteca	No
Accesos a la página de la biblioteca	No
Acceso al catálogo de la biblioteca	No
Acceso a servicio de descubrimiento (PRIMO)	No
Acceso a la plataforma de servicios bibliotecarios	No
Acceso a las bases de datos	No
Análisis de proxys	No
Dirección IP del host de acceso	No
Dominio (Dirección http, https y/o www)	No
Recurso electrónico empleado	No
Título electrónico empleado	No
Interacciones en RRSS de la biblioteca	No
Facebook	No
Twitter	No
Instagram	No
Blog	No
WhatsApp	No
Análisis del uso de las instalaciones bibliotecarias	No

### ¿Por qué o para qué se efectuó la transacción?

Fines para los cuales se efectuó la transacción

<b>VARIABLES DESEABLES</b> <b>Motivo para realizar la transacción</b>	<b>VARIABLES DISPONIBLES BDESSalud</b>
Seguimiento a transacciones	No
Encuestas a usuarios	No
Entrevistas a usuarios	No
Grupos focales	No
Determinar el impacto académico del uso de los recursos	No
Análisis de citas (producción académica de los usuarios)	No
Análisis bibliométricos (producción académica de los usuarios)	No

Fuente: Elaboración propia.

## Anexo 7. Políticas de carga para el llenado del DW BDESSalud

Las siguientes convenciones fueron determinadas con el fin de integrar de manera coherente tanto los datos provenientes de la fuente de información transaccional como de fuentes ajenas a ésta, de manera que no se pierda información sobre los movimientos de circulación de MDoc realizados y así mantener la **calidad, coherencia e integridad** de los datos contenidos en el DW.

- En el caso del usuario con valores en el campo substr303 que no encuentre coincidencia en las tablas Z303-Z305 y Z308 (ya sea porque maneja valores alfanuméricos o simplemente no se encuentre en la tabla referida) será señalado en el DW como “Desconocido”. Esto se hace con el fin de no perder el registro de la transacción realizada.
- En el caso del bibliotecario que aparezca con el campo de identificación en blanco, serán señalado como “Desconocido”. Esto se hace con el fin de no perder el registro de la transacción realizada.
- La transacción que no esté ligada a algún registro válido de MDoc manejará en el campo referido a éste el texto “MDoc no especificado”.
- El ítem (Z30H) que no esté ligado a algún registro válido de MDoc (Z13) manejará en el campo referido a éste el texto “MDoc no especificado”.
- Todas aquellas celdas vacías, que solo contengan espacios, o que contengan específicamente el texto “[NULO]” o alguna de sus variantes, son ingresadas al modelo transaccional con valores de ‘NULL’.
- Dependiendo de la funcionalidad esperada de la variable y en caso de que ésta no cumpla con los criterios necesarios para llevarla a cabo, se realizará su debida adaptación para que cumpla con los criterios para ser ingresada. En caso de no contar con la información suficiente para solucionar la inclusión de tales datos en específico, éstos serán excluidos de la inclusión a las tablas transaccionales, siempre buscando la menor perdida de información posible.
- En el caso de las tablas del DW que estén conformadas por valores obtenidos de distintas tablas transaccionales, las cuales se relacionan por la llave primaria, existe la posibilidad de que se haga referencia a ítems que no se encuentren en las demás, ocasionando variables incompletas. En estos casos se agregarán los valores al DW con el texto NULL en los campos de aquellas columnas con las que no se encuentre información.



## **Anexo 8. Políticas y estrategias de actualización del DW BDESSalud**

Las siguientes convenciones fueron determinadas con el fin de darle continuidad a la integración coherente de datos que busquen enriquecer el planteamiento inicial del DW BDESSalud establecido en este proyecto. Aquí es necesario enfatizar que siempre deben estudiarse los cambios que puedan producirse en las fuentes de datos, de manera que se tenga la noción de si dichos cambios llegan a afectar el contenido del DW y al proceso de actualización de este. En caso de que no haya habido cambios o que los producidos no afecten de manera significativa al DW, se considerarán las siguientes recomendaciones:

- La información se actualizará de manera semestral.
- Los datos de las tablas de Subdimensiones TIPO\_USUARIO, CARRERA y UNIDAD ACADÉMICA, serán cargadas siempre en su totalidad. Esto será siempre y cuando no hayan ocurrido cambios en las fuentes de información que sustentan dichas Subdimensiones. En caso de que haya habido cambios importantes, se deberán llevar a cabo las actualizaciones pertinentes en el DW para darles cabida a los nuevos datos.
- Los datos de las tablas de Dimensiones MDOC, USUARIO, TIPO\_TRANSACCION, TIPO\_DEVOLUCION, BIBLIOTECARIO y MULTA se cargarán de forma acumulativa considerando la creación de nuevos registros de las tablas que las alimentan a partir de la fecha de la última actualización.
- Los datos de la tabla Dimensión TIEMPO se cargarán de forma incremental teniendo en cuenta la fecha de la última actualización.
- El proceso *ETL* para la actualización del DW es similar al de carga inicial.

## **Anexo 9. Catálogo de códigos para valores de las tablas Oracle Aleph del SUBA**

### **TIPO DE MATERIAL**

Desglose de los códigos alfanuméricos para la columna Z30H\_MATERIAL. Coincide con la Z30\_MATERIAL y la Z35\_MATERIAL.

AUDIO= Audio  
CDROM= Disco compacto  
DIGIT= MDoc digital  
DISCO= Disco  
DISK= Disco  
DOCUM= Documento  
DVD= DVD  
ISSBD= ISSBD  
ISSUE= ISSUE  
LIBRO= Libro  
MAPA= Mapa  
PART= Partitura  
REVIS= Revista  
VIDEO= Video  
(Vacías)= No identificado

### **TIPO DE COLECCIÓN**

Desglose de los códigos alfanuméricos para la columna Z30H\_COLLECTION. Válido también para la columna Z30\_COLLECTION.

ACERV= Acervo general  
COLES= Colecciones especiales  
CONSU= Consulta  
ELECT= Medios electrónicos  
FOTOC= Fotocopias  
LECTU= Lecturas  
NOVEL= Novelas  
PP= Publicaciones periódicas  
RESER= Reserva  
TESIS= Tesis  
BIOGR= Biografías  
CIATE= Colección CIATEC  
MAEST= Maestría  
MEMOR= Memoria  
ESPEC= Programa especial  
MUSIC= Música  
MANUS= Manuscritos

CARTO= Cartografía

### **TIPO DE ADQUISICIÓN**

Desglose de los códigos numéricos para la columna Z30H\_ITEM\_STATISTIC. Es equivalente a la columna Z30\_ITEM\_STATISTIC.

00= Anteriores  
01= Compra directa  
02= Proyecto institucional  
03= Donación por otra institución  
04= Donación por titulación  
05= Donación por multa  
06= Reposición por pérdida  
07= Reposición por material dañado  
08= Canje  
09= Otros

IMPORTANTE: Para habilitar el uso de estos códigos se tiene que agregar un cero a los códigos de la Tabla de Oracle de la columna Z30H\_ITEM\_STATISTIC, pues deben ser de dos dígitos para encontrar su correspondencia con los valores de arriba.

### **ESTADO DEL MDOC**

Empleado para definir el tipo de préstamo. Válidos para las columnas Z30H\_ITEM\_STATUS, Z30\_ITEM\_STATUS y Z35\_ITEM\_STATUS.

Desglose de los códigos numéricos para la columna Z30H\_ITEM\_STATUS

01=3 días externo  
02=Préstamo interno  
03=1 día externo  
04=15 días externo  
05=2 horas  
06=Cubículo  
11=PIB  
90=Adquisiciones  
91=Encuadernación  
92=Perdido  
93=En proceso

### **CAMBIOS HECHOS AL MDOC**

Desglose de las opciones alfanuméricas para la columna Z30H\_H\_REASON

ITEM-LOCATION= Ubicación  
ITEM-LOCATION-2= Ubicación 2  
ITEM-COLLECTION= Colección  
ITEM-SUB-LIBRARY= Sub biblioteca  
ITEM-PROCESS-STATUS= Estatus proceso  
ITEM-BARCODE= Código de barras  
CREATE= Crear  
BIND= Encuadernar  
DELETE= Eliminado  
ITEM-STATUS= Estatus  
ALL-FIELDS= Todos los campos

### **ÚLTIMO CAMBIO HECHO AL MDOC**

Desglose de los códigos alfanuméricos para la columna Z30H\_H\_REASON\_TYPE

CHANGE= Cambio  
DELETE= Eliminado

### **ESTADO DE LA MULTA**

Desglose de los códigos numéricos para la columna Z31\_STATUS

C= Pagada  
O= Pendiente  
W= Cancelada

### **TIPO DE MULTA**

Desglose de los códigos numéricos para la columna Z31\_TYPE

0000 - All  
0001 - Photo request  
0002 - Hold request  
0003 - Late return  
0004 - Loan  
0005 - Renewal  
0006 - Photocopy request processing  
0007 - Photocopy request not filled  
0008 - Photo. request home delivery  
0009 - Photocopy request pickup  
0010 - Claim return  
0011 - ILL request  
0013 - Routing list  
0014 - Late return of routed issue

0015 - ILL material arrival  
0016 - Incoming ILL request  
0017 - Issuing library card  
0021 - Borrower registration  
0022 - Borrower renewal  
0023 - New user  
0024 - Hold request filled - 01  
0025 - Hold request not filled  
0026 - Hold request filled - 02  
0027 - Hold request filled - 03  
0028 - Hold request filled - 04  
0040 - Lost material - Handling  
0041 - Lost material - Replacement  
0042 - Lost material - Notice  
0050 - Recall late return fine  
0051 - Rush recall late return fine  
0052 - Recall fine limit  
0053 - Rush recall fine limit  
0054 - Recall lost letter  
0055 - Rush recall lost letter  
0080 - 1st warning - Overdue  
0081 - 2nd warning - Overdue  
0082 - 3rd warning - Overdue  
0083 - 4th warning - Overdue  
0090 - Overdue summary  
0091 - Booking Request  
0092 - Booking Request Filled  
0093 - Booking Request Delivered  
0094 - Booking request not filled  
1024 - Hold request filled - 01  
1026 - Hold request filled - 02  
1027 - Hold request filled - 03  
1028 - Hold request filled - 04  
9000 - Deposit  
9001 - Return deposit  
9995 - Use of copyright restricted object  
9996 - Local charge for disruption  
9997 - Damaged material  
9998 - Donation  
9999 - Payment

### **TIPO DE TRANSACCIÓN**

Desglose de los códigos numéricos para la columna Z35\_EVENT\_TYPE

50=Préstamo externo  
52=Circulación fuera de línea  
58=Préstamo eliminado  
61=Devolución  
62=Renovación - Usuario (Web)  
63=Renovación - Bibliotecario  
65=MDoc declarado perdido  
66=MDoc reclamado devuelto  
71=Reservación  
73=Solicitud de reserva  
74=Retener transacción de solicitud (colocada en el MDoc no disponible, sin expansión)  
80=Préstamo interno  
82=Devolución de MDoc no prestado (Error)  
100=Multas

### **TIPO DE ORIGEN DE LA TRANSACCIÓN**

Desglose de los códigos alfanuméricos para la columna Z35\_TYPE

P= PC\_Server  
W= Web

### **ESTADO DEL PRÉSTAMO**

Desglose de los códigos numéricos para la columna Z36H\_STATUS. Es equivalente a la columna Z36\_STATUS.

A=Préstamo activo  
C=Reclamación del MDoc (Devolución)  
L=MDoc perdido

### **BIBLIOTECA DE ADSCRIPCIÓN**

Desglose de los códigos alfanuméricos para la columna Z303\_HOME\_LIBRARY

DS=DES DE SALUD

## **TIPO DE USUARIO**

Desglose de los códigos numéricos para la columna Z305\_BOR\_STATUS

- 01= Alumno
- 02= Académico
- 03= Administrativo
- 04= Externo
- 05= Préstamo interno
- 06= Posgrado
- 07= Sistema abierto

## **Anexo 10. Política de seguridad de datos del SUBA-UACH**

Estimado usuario:

La información referente a los datos personales de identificación (nombre, domicilio, teléfonos y correo electrónico) que usted proporciona a la biblioteca o al Sistema Universitario de Bibliotecas Académicas, está resguardada por quien la maneja para evitar el uso indebido por parte de personal no autorizado.

Toda la información proporcionada es confidencial y privada en cualquiera de sus formatos impresos o electrónicos.

El SUBA recaba sus datos personales sólo cuando usted los proporciona directamente en las bibliotecas o en el momento de inscribirse como estudiante a través de la Dirección Académica.

Sus datos personales son confidenciales y solo serán utilizados para otorgar servicios académicos como el préstamo de material documental o para informarle sobre nuevas adquisiciones de libros.

El SUBA se compromete a que sus datos serán utilizados bajo las más estrictas medidas de seguridad que garanticen su confiabilidad permaneciendo siempre protegidos con contraseña en nuestro sistema ¡Mediante nuestra política de seguridad se garantiza la integridad de la información proporcionada por usted!

Fuente: Página web del SUBA-UACH (<https://uach.mx/suba/>).



## **Anexo 11. Temáticas para considerar en la formación de los bibliotecarios de datos**

De acuerdo con la investigación de Lyon et al (2015) referente al desarrollo curricular en la Ciencia de datos, los cursos que tendrían que considerarse en la formación de los bibliotecarios de datos serían los siguientes:

<b>Esenciales</b>	<b>Deseables</b>
<ul style="list-style-type: none"><li>• Gestión de datos</li><li>• Infraestructuras de datos</li><li>• Metadatos</li><li>• Bibliotecas académicas</li><li>• Preservación de colecciones digitales</li><li>• Metodologías de investigación</li></ul>	<ul style="list-style-type: none"><li>• Introducción a las tecnologías de la información</li><li>• Administración y dirección de servicios de información</li><li>• Repositorios digitales</li><li>• Sistemas de Información Geográfica para bibliotecarios</li><li>• Visualización de información</li><li>• Programación para bibliotecarios</li><li>• Introducción al análisis estadístico de datos</li></ul>

---

Fuente: Traducción propia de la propuesta de Lyon et. al. (2015).

## Anexo 12. Identificación de iniciativas de aplicación de metodologías de la Ciencia de datos en bibliotecas mexicanas

Año	Autor/Institución	Propuesta (Modelo creado, software empleado)	Procesos en que incide	Formación del autor
2016	Germán de Jesús Hernández García / Subdirección del Sistema de Información Bibliotecario de Ecosur	Uso de herramientas ETL a partir de tecnologías Aleph. Herramienta que permite la extracción, transformación y carga de datos a los distintos sistemas que se utilizan en el Sistema de Información Bibliotecario de ECOSUR y a otras plataformas de terceros en las que colabora la biblioteca / Software: Uso de API's de Aleph y del software Mongo DB.	Extracción de datos (ETL) para repositorio institucional (tesis) Mongo DB. Acceso abierto. Extracción de metadatos bibliográficos del catálogo, transformarlos con los metadatos solicitados en el repositorio institucional y se cargaron en el RI.	Técnico en Computación Licenciatura en Ingeniería en Sistemas Computacionales
2018	Ing. Federico Portas-Lagar / Jubilado del ITESO	Servicio ITMS Analytics / Software no especificado	Se adapta a las necesidades de las bibliotecas que contratan el servicio, se especializa en los temas de presupuestos, utilización de servicios y tendencias de uso.	Ingeniería (no especificada)
2018	Nicholas Cop / DiS Soluciones Innovadoras	Big Data para bibliotecarios académicos / Software: Tableau Public, Open Refine, etc.	Conceptos teóricos generales respecto al tema de Big Data y realización de prácticas sencillas de limpieza y visualización de datos.	Licenciatura en Física. Estudios de posgrado en Física Nuclear Experimental.
2018	Mtra. Carolina Gutiérrez / FM UNAM	Aplicación de MD al proceso de circulación del material de la biblioteca de la Facultad de Medicina de la UNAM / Software: Excel, Weka y Rapid Miner	Proceso de circulación bibliográfica.	Licenciatura en Informática Maestría en Ciencias de la Informática

<b>Año</b>	<b>Autor/Institución</b>	<b>Propuesta (Modelo creado, software empleado)</b>	<b>Procesos en que incide</b>	<b>Formación del autor</b>
2018	Mtro. Bernardo Ramírez Lujano / UAQ	Visualización de datos como apoyo para la toma de decisiones del Sistema Bibliotecario Universitario de la Universidad Autónoma de Querétaro / Software: Excel y Tableau Public.	Circulación de material documental. Préstamos. Toma de decisiones en Desarrollo de Colecciones. Corrección de errores en catalogación.	Psicólogo Social Maestría en Diseño e Innovación UAQ
2018	Dr. Raymundo Miguel Campos Vázquez / BDCV COLMEX	Proyecto sobre la circulación de materiales de la BDCV. Entrega tardía de materiales y relación con tipo de usuario, centro de adscripción, carrera, tema y promedio académico / Software no especificado	Circulación de material documental.	Licenciatura en Economía ITESM Maestría en Economía COLMEX Doctorado en Economía UC Berkeley
2018	Dr. Marcial Contreras Barrera / DGB UNAM	Aplicación de la minería de texto a los registros de la base de datos del catálogo LIBRUNAM / Software: Lenguaje de programación PHP y manejador de bases de datos MySQL. Utilizó el método de minería de texto RAKE ( <i>Rapid Automatic Keyword Extraction</i> ).	Diseño de un modelo de minería de texto que apoya la labor de organizar la información documental.	Licenciatura en Ingeniería en Computación Maestría en Ingeniería Eléctrica Doctorado en Bibliotecología y Estudios de la Información
2018	M. C. Laura Ortigoza / INFOESTRATÉGICA	Intelligent / Software no especificado	Ofrece información sobre tendencias de uso, frecuencia e interés en los recursos documentales, así como el comportamiento de los usuarios. eContent, COUNTER, OPAC y Circulación.	Ingeniería Química UV Maestría en Ciencias CINVESTAV

Fuente: Elaboración propia.