

Análisis comparativo de herramientas para la gestión de metadatos

Comparative analysis of tools for metadata management

Mike Castilla, José Lora, Raúl Vega, & Waldir Portilla

[\[dianahv@unisimonbolivar.edu.co\]](mailto:dianahv@unisimonbolivar.edu.co)

Universidad Simón Bolívar, Barranquilla-Colombia

Resumen Muchas de las organizaciones no saben elegir al momento de adquirir una herramienta para la gestión de metadatos. Es debido a esto que uno de los objetivos principales de este estudio es conocer algunas de las herramientas de gestión de metadatos que existen actualmente en el mercado. Además, se desea comparar algunas de estas herramientas haciendo pruebas de inserción de datos y medir su rendimiento para así elegir la más eficiente y sencilla de manejar.

Pero para que sea más fácil la comprensión de estas herramientas primero definiremos ¿Qué son los metadatos y la gestión de metadatos? Los metadatos son “datos acerca de los datos” y sirven para suministrar información sobre los datos producidos. Los metadatos consisten en información que caracteriza datos, describen el contenido, calidad, condiciones, historia, disponibilidad y otras características de los datos. La gestión de metadatos no es más que el control que se tienen de estos metadatos

Palabras clave: Metadatos, Gestión de Metadatos, Gobernabilidad de datos

Abstract Many of the organizations do not know how to choose when acquiring a metadata management tool. It is because of this that one of the main objectives of this study is to know some of the metadata management tools that currently exist in the market. In addition, you want to compare some of these tools by testing data insertion and measuring their performance to choose the most efficient and easy to handle.

But to make it easier to understand these tools we will first define what are metadata and metadata management? Metadata is "data about data" and serves to provide information on the data produced. Metadata consists of information that characterizes data, describes the content, quality, conditions, history, availability, and other characteristics of the data. Metadata management is only the control you have of these metadata

Key words: Metadata, Metadata Management, Data Governance.

Introducción

la importancia de los metadatos dentro del gobierno de datos es debido a que proporcionan una organización muy clara acerca de los datos. Dentro de la práctica de gobierno de datos, interpretar el contexto del contenido de los datos es esencial para administrarlos de la mejor manera y los metadatos son la clave para hacerlo, de ahí su importancia.

Los metadatos son “datos acerca de los datos” y sirven para suministrar información sobre los datos producidos. Los metadatos consisten en información que caracteriza datos, describen el contenido, calidad, condiciones, historia, disponibilidad y otras características de los datos. Con base en los metadatos se manejan herramientas que sirven para gestionar o administrar los datos que se encuentran suministrados. [1]

El control (gestión) de los metadatos es indispensable y, por este motivo, se hace necesario disponer de mecanismos que aporten visibilidad sobre el ciclo del dato. La gestión de metadatos ayuda a alcanzar este objetivo, contextualizando y aportando un sentido a toda la información contenida en el sistema. De este modo se eleva el valor comercial de los activos de datos del negocio y se puede responder a cuestiones acerca del significado de un término, las políticas y reglas que le son de aplicación. [2]

De allí surge una incógnita ¿cómo gobernar los metadatos de manera, segura, organizada, y menos compleja?

En el mercado se pueden encontrar gran cantidad de herramientas para gestión de metadatos, con diversas características y funcionalidades, las cuales pueden ser utilizadas por empresas u organizaciones que requieran implementarlas. Pero, surgen varias preguntas: ¿Qué tipo de herramienta sería la más adecuada para una organización, según sus características?, ¿Es fácil de utilizar?, ¿Cuál es su costo?, ¿A qué tipo de objetos de datos se pueden registrar sus metadatos? ¿Qué información se debe registrar como metadatos? entre otras.

El objetivo principal del presente proyecto es comparar alguna de las herramientas de gestión de metadatos las cuales se encuentren en el mercado, para analizar sus características y sus beneficios. Para lograr este objetivo se evaluará el funcionamiento y uso de estas herramientas de acuerdo a sus características.

Este proyecto busca dar a conocer tanto a entidades públicas como privadas qué herramientas para la gestión

de metadatos puede brindar un primer acercamiento a ellos para su posterior implementación.

Estado del arte

Constantemente están surgiendo nuevas herramientas para la gestión de metadatos, las cuales nos sirven para ampliar la búsqueda de información de un documento.

Una de las empresas que ofrece entre sus servicios la gestión de metadatos se llama **Capgemini**, una empresa francesa, que a través esta actividad garantiza el valor continuo de los activos de datos de una empresa. La solución de gestión de metadatos que ofrecen permite la integración con portales de metadatos, obtención y extracción de metadatos de diferentes fuentes para facilitar su actualización, mantenimiento, integración, depuración y consolidación, entre muchas otras funcionalidades. [3]

Otra de estas herramientas es **Talend Metadata Manager** que conecta datos desde plataformas, bases de datos y herramientas analíticas para generar una visión integral de la cadena de suministro de información en un lenguaje comprensible para todos. Permite registrar metadatos a cada objeto de datos utilizado en una tarea de análisis de datos. [4]

Collibra también es una herramienta de gestión de metadatos que facilita que los ciudadanos de datos encuentren, entiendan y confíen en los datos de la organización que necesitan para tomar decisiones comerciales todos los días. A diferencia de las soluciones de gobierno de datos tradicionales, Collibra es una plataforma interorganizacional que descompone los silos de datos tradicionales, liberando los datos para que todos los usuarios tengan acceso. Brinda este tipo de productos: Glosario empresarial, Diccionario de datos, Servicio de ayuda de datos, Gerente de política, Dato de referencia, Administración. En los cuales podemos solicitar una demostración de los productos mencionados. [5]

Oracle Enterprise Metadata Management también es una herramienta de gestión de metadatos. Mejora la búsqueda de metadatos a través de búsquedas de texto y capacidades de exploración de metadatos proporcionando el linaje de datos e informes de análisis de impacto en todas las tecnologías.

También Garantiza la estandarización de datos y la gestión de datos a través de la gestión de metadatos

aumentando a la efectividad de la mayordomía a través de una mayor colaboración, características sociales y el cumplimiento de los datos a través del control de versiones y los informes de metadatos.

[6]

Metadata Manager:

Metadata Manager conecta todos los metadatos desde plataformas de inteligencia corporativa, de las bases de datos y desde herramientas de modelación hasta analíticas para formar una cadena de suministro visual de la información de la organización. Acelera sus proyectos de gestión del cambio, migración y modificación de plataformas con una visión completa de los flujos de datos y un análisis de impacto. Metadata Manager recaba, conecta e integra metadatos en distintas plataformas y permite visualizar el impacto que puede causar cualquier cambio antes de propagarlo por todo su entorno de datos.

[7]

IBM también ofrece una solución de gestión de metadatos. InfoSphere Information Server de IBM presenta un repositorio de metadatos que almacena los metadatos de las herramientas de la suite y las herramientas y bases de datos externas y permite el intercambio entre ellos. Los usuarios pueden importar metadatos en el repositorio desde múltiples fuentes, exportar metadatos por varios métodos y transferir metadatos entre los repositorios de diseño, prueba y producción. Los cambios que se realizan en el repositorio se realizan automáticamente en todo el conjunto y utilizan tecnología de base de datos relacional estándar.

[8]

Alex Solutions:

Alex Solutions es un mercado de datos empresariales que permite obtener resultados comerciales centrados en datos. Reconocido por Gartner y Forrester como líder en el suministro de soluciones en dominios de Detección de datos y Calidad de datos Perfilado, Gestión de la gobernanza de la información, Gestión de riesgos y Gestión de metadatos. [9]

Descripción del trabajo realizado

Definición de parámetros a evaluar

Para evaluar las herramientas de gestión de metadatos se definieron una serie de parámetros o características. En las

pruebas realizadas con cada herramienta se buscó obtener una medición de cada parámetro.

Los parámetros elegidos fueron:

- **Facilidad de uso:** Interfaz amigable con el usuario, fácil acceso, fácil comprensión de los datos requeridos, fácil visualización de los resultados.
- **Idioma:** Si la herramienta es multi-idioma, configurable, teniendo en cuenta que la herramienta debería ser utilizable por personas en cualquier parte del mundo.
- **Ítem de metadatos que solicita la herramienta:** Cada uno de los elementos que solicita registrar. Esto para determinar la completitud de metadatos.
- **Multi-usuario:** Si la herramienta permite el trabajo colaborativo entre varios usuarios.
- **Cantidad de objetos a los cuales se les puede registrar metadatos:** Si la herramienta tiene una cantidad límite o máxima.
- **Tipos de objetos de datos a los cuales se les puede registrar metadatos:** Si la herramienta es flexible en el tipo de objetos.
- **Obligatoriedad de los campos de metadatos:** Si hay campos de metadatos obligatorios o no; también puede ser indicador de flexibilidad.
- **Valores específicos para los campos de metadatos:** Si dentro de los campos solicitados, hay algunos que exijan un tipo o valor específico o son totalmente abiertos. También es indicador de flexibilidad.
- **Registro de metadatos de la data set separado del registro de metadatos de los atributos:** Si se pueden administrar los metadatos por separado.
- **Facilidad de actualización/modificación de metadatos:** Facilidad de gestión, sin pérdida de datos, sin alteración de metadatos previamente registrados.

Herramientas utilizadas

En el transcurso de la investigación se hizo el intento de obtener algunas versiones de prueba (demo) de herramientas propietarias tales como Collibra, Alation, Alex Solutions, a través de sus sitios web, en los cuales no

hay opción de descarga directa de dichas versiones, sino una solicitud a la empresa. No se logró respuesta en ninguna de estas solicitudes, por lo cual se optó por buscar herramientas con licencia GNU o libres. De esta forma se obtuvieron Talend [10] y Colectica [11] en su versión para Excel.

Pruebas de registro de metadatos

En Colectica para Excel:

Como pruebas se insertó un libro de datos, para luego llenar cada ítem que colectica propone llenar. Cabe destacar que la inserción de metadatos es muy sencilla.

Colectica pidió una serie de datos a llenar, de los cuales llegamos a la conclusión de que no hay campos obligatorios. Como default Colectica nos pide en su pestaña de “detalles de conjunto de datos (en español)” un título, un subtítulo, un título alternativo, un creador, un editor, un contribuyente, derechos de autor y derechos de accesos. Hay un ícono que puede visualizar “+”, que sirve para agregar un campo adicional y darle el nombre que nosotros queramos, y así detallar más la documentación.

Se quiso agregar información detallada sobre una columna en su pestaña “Detalles de variable” que solicita los siguientes metadatos: etiqueta, descripción, tipo de datos, que tiene un selector de ítem, y por ende unos valores predeterminados a escoger tales como texto, numérico, tiempo de datos, código. Aditividad, es otro campo que también tiene un selector predeterminado, y estos son sus ítems: Sin especificar, valores, flujo de datos, no aditivo [39-42].

Otros de los campos que pide es, unidad de medida, rol, unidad de análisis, unidad de respuesta.

Da la posibilidad de agregar un campo extra, por si quiere detallar más los metadatos de la columna.

Concluimos también lo mismo que la anterior pestaña (detalles del conjunto de datos), que no pide campos obligatorios a llenar, pero como obviedad contiene más campos default.

Nota: Esta herramienta permite documentar libros de datos.

En Talend Open Studio

Para añadir metadatos en el programa de TALEND OPEN STUDIO se puede registrar metadatos a archivos en los siguientes formatos XML, EXCEL, JSON, entre otros.

Para ingresar metadatos en el tipo de formato seleccionada le damos clic derecho y nos permitirá crear archivos (XML, EXCEL, JSON), crear carpetas, importar y exportar ítems

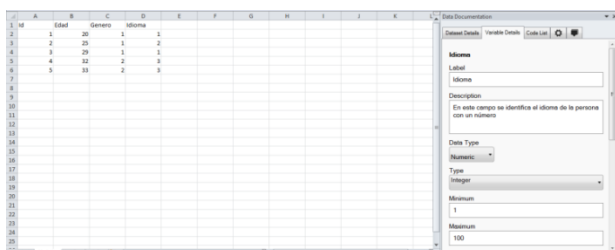
Al añadir un metadato pedirá un nombre, descripción, estados (desarrollo, pruebas o producción) luego de esto se selecciona si los metadatos son de entrada o de salida, se configura los parámetros del archivo, este muestra la tabla en la cual están guardados los metadatos

la herramienta también permite importar archivos los cuales al momento de ingresarlos y cargarlos genera una vista previa de los datos, el paso siguiente es definirles los parámetros, ya en el esquema final se tienen las columnas , marcar si es clave, se define el tipo (String, integer, boolean, byte, carácter, double, date, float, object, long, list, document) , se define si es nulo o si contiene algún dato , y con esto ya se tiene insertado los datos.

Para visualizar los datos insertados en la parte de los formatos, se puede desplegar los metadatos insertados y las columnas ingresadas.

Resultados

En el siguiente cuadro se establecieron algunos parámetros con las cuales se realizó la comparación entre las dos herramientas de gestión de metadatos. A través de dicho cuadro se puede observar qué herramienta es más eficiente y fácil de usar para este fin.



Cuadro comparativo:

Métricas	Herramientas	
	Colectica Excel	Talend open studio
1. Facilidad de uso	Su interfaz es amigable con el usuario, por ende se agrega información detallada de cada columna, agregar campos adicionales, etc. de una manera muy sencilla.	La interfaz es muy completa en la cual se encuentra de forma detallada los campos y formatos que se pueden utilizar.
2. Idioma	No contiene diferentes idiomas a la cual se pueda traducir, solo se encuentra disponible el inglés.	Se encuentra disponible en idioma inglés.
3. Parámetros que solicita para cada registro.	Label, description, data type, type, minimum, scale, decimal position, addivity, measurement unit, role, analysis unit, response unit	Name, purpose, description, state, parameters, type, role, clave, file, position.
4. Herramienta multiusuario.	La herramienta solo permite acceso de un usuario. Por eso se le considera que no es multiusuario.	La herramienta multiusuario se encuentra habilitada en la versión licenciada, la versión demo solo permite un usuario.
5. Límite de objetos a los cuales se les puede registrar metadatos	No definido	No definido
6. Qué tipo de objetos/documentos permite registrar metadatos.	Permite registrar metadatos para cada libro de Excel. Dentro de este...	Permite registrar XML, EXCEL, JSON, SQL, REGEX, LDIF
7. Obligatoriedad de los campos de metadatos	Ninguno es obligatorio	Casi todos, son pocos los que no son obligatorios.
8. Valores específicos para los campos de metadatos	Muchos son de formato libre, pero otros tienen una lista de valores predefinidos.	Muchos son de formato libre, pero otros tienen una lista de valores predefinidos.
9. Registro de metadatos para el data set (objeto de datos) separado de los metadatos de los atributos.	Si, en formatos diferentes, además cada atributo solicita sus metadatos por separado.	Si, en los formatos antes mencionados, cada cual tiene sus metadatos en espacios diferentes.
10. Facilidad de actualización /	Fácilmente modificables. Se puede modificar cada campo por separado.	Fácilmente modificables.

modificación de metadatos.		
----------------------------	--	--

Conclusiones

Las herramientas de metadatos se hacen necesarias para disponer de mecanismos que aporten visibilidad sobre el ciclo del dato. Por este motivo con el desarrollo de este proyecto se pretende dar las maneras y razones a la comunidad interesada en el tema de escoger una herramienta de gestión de metadatos tomando como punto de inclinación la más eficiente y sencilla.

La obtención de Colectica fue muy sencilla basta de entrar a la página oficial y seguir unos pequeños pasos, entre ellos hacer un registro y de descargará un archivo que es el instalador de la herramienta. Colectica tiene una facilidad de uso impresionante, es sencilla y versátil.

Es debido a la facilidad de uso de Colectica Excel que se les recomienda a los usuarios principiantes comenzar con esta herramienta, o bien pueden usar directamente Talend Open Studio ya que ambas se le considera de fácil acceso, ésta es una de las similitudes que presentan estas dos.

Ambas herramientas proporcionan algunas diferencias. En el caso de Talend Open Studio se le considera una herramienta multiusuario en su versión licenciada, a diferencia de Colectica Excel que solo mantiene una versión la cual no es multiusuario.

Una de las desventajas de usar Colectica Excel es que permite registrar metadatos solamente para los libros de Excel. En cambio, Talend Open Studio permite registrar para XML, EXCEL, JSON, SQL, entre otras.

Referencias

- [1] Digital, Secretaria de Gobierno, «geoidep,» Presidencia de consejo de ministros, [En línea]. Available: <https://www.geoidep.gob.pe/conoce-las-ides/metadatos/que-son-los-metadatos>. [Último acceso: 27 Mayo 2019].
- [2] Power Data, «blog.powerdata,» [En línea]. Available: <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/353207/la-gesti-n-de-metadatos>. [Último acceso: 2019 Mayo 27].
- [3] Capgemini, «Capgemini,» Capgemini, [En línea]. Available: <https://www.capgemini.com/>. [Último acceso: 12 Febrero 2018].

- [4] Talend, «es.talend,» Talend, 2006. [En línea]. Available: <https://es.talend.com/>. [Último acceso: 27 Mayo 2019].
- [5] Collibra, «Collibra,» Collibra, [En línea]. Available: <https://www.collibra.com>. [Último acceso: 27 Mayo 2019].
- [6] Oracle, «Oracle Enterprise Metadata Management,» Oracle, [En línea]. Available: <https://www.oracle.com/middleware/technologies/enterprise-metadata-management.html>. [Último acceso: 27 Mayo 2019].
- [7] Talend, «Metadata Manager,» Talend, [En línea]. Available: <https://es.talend.com/products/data-catalog/>. [Último acceso: 27 Mayo 2019].
- [8] IBM, «InfoSphere Information Server,» IBM, [En línea]. Available: <https://www.ibm.com/cos/analytics/information-server>. [Último acceso: 27 Mayo 2019].
- [9] Alex Solutions, «Alex solutions,» [En línea]. Available: <http://alexolutions.com.au>. [Último acceso: 27 Mayo 2019].
- [10] Talend, 2005.
- [11] Colectica, «Colectica,» [En línea]. Available: <https://www.colectica.com/software/colecticaforexcel/>.
- [12] Greenpeace, «www.greenpeace.org,» [En línea]. Available: <http://www.greenpeace.org/colombia/es/campanas/cambio-climatico/>. [Último acceso: 27 04 2016].
- [13] J. S. C. Papaleo, «DW,» [En línea]. Available: <http://www.dw.com/es/el-protocolo-de-kioto-fracas%C3%B3-hace-tiempo/a-15600811>. [Último acceso: 27 04 2016].
- [14] OMM - Organización Meteorológica Mundial, Cambio Climático 2014. Informe de Síntesis, Ginebra: OMM, 2015.
- [15] A. Morales, M. C. Escoto, R. Garcia, J. Molinar y C. Hidalgo, «Sistema para la Aplicación de Pruebas Psicológicas via Web,» *Acta Universitaria*, vol. 22, Abril 2012.
- [16] J. Molinar, M. d. C. Escoto, R. Z. García y E. Bautista, «Evaluación Computarizada De Pruebas Psicológicas Mediante El Procesamiento Digital De Imágenes,» *Enseñanza e Investigación en Psicología*, vol. VOL 17, nº 2: 415-426, pp. 1 - 14, Diciembre 2012.
- [17] R. J. Gregory, Pruebas Psicológicas: Historia, principios y aplicaciones, Sexta ed., Pearson, 2012.
- [18] K. F. Geisinger, «Psychological Testing of Hispanics,» *APA Science Volume Series*, Junio 1992.
- [19] C. Santiesteban, Principios De Psicometria, SINTESIS PSICOLOGIA, 2009.
- [20] M. Sanchez, «Responsabilidad Científica Del Psicólogo,» *Psicogente*, pp. 155-168, 2013.
- [21] J. Cohen R, Pruebas y evaluación Psicológicas, McGraw-Hill, 2001.
- [22] R. Jay, Pruebas y evaluación psicológica : introducción a las pruebas y a la medición., Segunda ed., McGraw-Hill, 2006.
- [23] J. H. Ávila, S. L. Hoyos, D. P. González y A. Cabrales, «Relación Entre Ansiedad Ante Los Exámenes, Tipos De Pruebas Y Rendimiento Académico En Estudiantes Universitarios,» *Psicogente*, vol. 14, pp. 256-268, Noviembre 2011.
- [24] «Definicion de Aplicacion Web,» [En línea]. Available: <http://www.alegsa.com.ar/Dic/aplicacion%20web.php>. [Último acceso: 9 Mayo 2014].
- [25] R. L.-. Pressman, Web Engineering. A Practitioner Approach, McGraw-Hill, 2010.
- [26] R. F. Grove, Web Based Application Development, Jones and Burtlett Publishers, 2008.
- [27] [En línea]. Available: <http://php.net/manual/es/intro-what-is.php>.
- [28] «Que es HTML5?,» [En línea]. Available: <http://www.immagen.com/Blog/Que-Es-HTML5>. [Último acceso: 22 Mayo 2014].

- [29] «MySQL The Most Popular Open-Source Database,» [En línea]. Available: <http://www.oracle.com/us/products/mysql/overview/index.html>. [Último acceso: 10 Mayo 2014].
- [30] «Netbeans - Overview,» [En línea]. Available: <https://netbeans.org/features/>. [Último acceso: 2 Mayo 2014].
- [31] «DBWrench,» [En línea]. Available: <http://www.dbwrench.com/index.shtml>. [Último acceso: 25 Abril 2014].
- [32] «Programación Extrema,» [En línea]. Available: <http://eisc.univalle.edu.co/materias/WWW/materia1/lecturas/xp.pdf>. [Último acceso: 1 Mayo 2014].
- [33] «Metodologías Ágiles En El Desarrollo Web,» [En línea]. Available: <http://blog.inteligencia.com/2007/01/metodologas-giles-en-el-desarrollo-web.html>. [Último acceso: 13 Mayo 2014].
- [34] J. H. Canós, P. Letelier y M. C. Penadés, «Metodologías Ágiles en el Desarrollo de Software,» *DSIC -Universidad Politécnica de Valencia*, pp. 1-8.
- [35] «Qué es MVC,» [En línea]. Available: <http://www.desarrolloweb.com/articulos/que-es-mvc.html>. [Último acceso: 25 Mayo 2014].
- [36] J. Framingham. [En línea]. Available: <http://psychcentral.com/lib/types-of-psychological-testing/0005924>. [Último acceso: 25 Abril 2014].
- [37] R. Kaplan y D. Saccuzzo, *Pruebas psicológicas: principios, aplicaciones y Temas*, Sexta ed., Cengage Learning Latin America, 2006.
- [38] *Capgemini*, 2018.
- [39] E. Steffens, D. Ojeda, O. Martínez, J. García, H. Hernández, 5; F. Marin., Niveles de pensamiento crítico en estudiantes de Universidades en Barranquilla (Colombia), *Espacios*, vol. 38 (30), 2017.
- [40] J. Ortega, *Importancia de la seguridad de los trabajadores en el cumplimiento de procesos, procedimientos y funciones*, Academia y Derecho, 14., 2017.
- [41] J. Jiménez, J.C. Mojica, H.G. Hernández., y Diego Cardona. Diagnóstico de la Innovación y Desarrollo Tecnológico en el Sector Hotelero de la Región Caribe Colombiana, *Información Tecnológica Vol. 29(5)*, 157-164, 2018.
- [42] J. D. Mojica Domínguez, A. Barandica Domínguez, L. Rodero Acosta, M. Franco Montenegro, H. Hernández Palma, S. Arboleda Cardona, La función administrativa en la era de las tic, *Revista Lasallista de Investigación*, vol. 12 (2), 2015.