Calidad de Datos

dqTeam

Ismael Caballero Muñoz-Reja • Ana Isabel Gómez Carretero
Fernando Gualo Cejudo • Jorge Merino García
Bibiano Rivas García • Mario Piattini Velthuis

Ra-Ma®

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos. RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:33:09.

Calidad de Datos
© Ismael Caballero Muñoz-Reja, Ana Isabel Gómez Carretero, Fernando Gualo Cejudo, Jorge Merino García, Bibiano Rivas García, Mario G. Piattini Velthuis
© De la edición: Ra-Ma 2018

MARCAS COMERCIALES. Las designaciones utilizadas por las empresas para distinguir sus productos (hardware, software, sistemas operativos, etc.) suelen ser marcas registradas. RA-MA ha intentado a lo largo de este libro distinguir las marcas comerciales de los términos descriptivos, siguiendo el estilo que utiliza el fabricante, sin intención de infringir la marca y solo en beneficio del propietario de la misma. Los datos de los ejemplos y pantallas son ficticios a no ser que se especifique lo contrario.

RA-MA es marca comercial registrada.

Se ha puesto el máximo empeño en ofrecer al lector una información completa y precisa. Sin embargo, RA-MA Editorial no asume ninguna responsabilidad derivada de su uso ni tampoco de cualquier violación de patentes ni otros derechos de terceras partes que pudieran ocurrir. Esta publicación tiene por objeto proporcionar unos conocimientos precisos y acreditados sobre el tema tratado. Su venta no supone para el editor ninguna forma de asistencia legal, administrativa o de ningún otro tipo. En caso de precisarse asesoría legal u otra forma de ayuda experta, deben buscarse los servicios de un profesional competente.

Reservados todos los derechos de publicación en cualquier idioma.

Según lo dispuesto en el Código Penal vigente, ninguna parte de este libro puede ser reproducida, grabada en sistema de almacenamiento o transmitida en forma alguna ni por cualquier procedimiento, ya sea electrónico, mecánico, reprográfico, magnético o cualquier otro sin autorización previa y por escrito de RA-MA; su contenido está protegido por la ley vigente, que establece penas de prisión y/o multas a quienes, intencionadamente, reprodujeren o plagiaren, en todo o en parte, una obra literaria, artística o científica.

Editado por:
RA-MA Editorial
Calle Jarama, 3A, Polígono Industrial Igarsa
28860 PARACUELLOS DE JARAMA, Madrid
Teléfono: 91 658 42 80
Fax: 91 662 81 39
Correo electrónico: *editorial@ra-ma.com*
Internet: *www.ra-ma.es* y *www.ra-ma.com*
ISBN: 978-84-9964-750-0
Depósito legal: M-21415-2018
Maquetación: Antonio García Tomé
Diseño de portada: Antonio García Tomé
Filmación e impresión: Safekat
Impreso en España en julio de 2018

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:41:35.

A Katty, in memoriam

Ismael Caballero Muñoz-Reja

A mis padres Ana Isabel y Conrado,
a toda mi familia y en especial a Bibi, que me dan
fuerza para embarcarme en cualquier aventura

Ana Isabel Gómez Carretero

A Fernando, Josefina y José Carlos,
mis padres y hermano, por su apoyo incondicional,
creer siempre en mí y hacerme mejor persona.

Fernando Gualo Cejudo

A mis amigos de DQTeam

Jorge Merino García

A mis padres, Bibiano y Loli, a mi familia,
amigos y Ana por caminar a mi lado.
Lo importante son los pasos, no el camino.

Bibiano Rivas García

A Óscar Díaz, por su amistad,
con admiración y gratitud.

Mario Piattini Velthuis

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:41:35.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:41:35.

ÍNDICE

AUTORES **11**

PRÓLOGO **15**

PREFACIO **19**

CONTENIDO 20

ORIENTACIÓN A LOS LECTORES 21

OTRAS OBRAS RELACIONADAS 22

AGRADECIMIENTOS 22

CAPÍTULO 1. FUNDAMENTOS DE LA CALIDAD DE DATOS **25**

1.1 CALIDAD DE DATOS EN EL MUNDO DIGITAL 25

1.1.1 Los datos y la transformación digital 25

1.1.2 Concepto de calidad de los datos 27

1.2 DATO, INFORMACIÓN, CONOCIMIENTO Y SABIDURÍA 28

1.3 CICLO DE VIDA DEL DATO 30

1.4 CAUSAS DE PROBLEMAS DERIVADOS DE NIVELES INADECUADOS DE CALIDAD DE DATOS 32

1.5 GOBIERNO VS. GESTIÓN DE DATOS 37

1.6 LECTURAS RECOMENDADAS 40

1.7 SITIOS WEB RECOMENDADOS 40

CAPÍTULO 2. DATOS MAESTROS (MASTER DATA) **41**

2.1 INTRODUCCIÓN A LOS DATOS MAESTROS 41

2.2 GESTIÓN DE DATOS MAESTROS 43

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:01.

2.3 RESOLUCIÓN DE ENTIDADES 45
2.4 ARQUITECTURA PARA MDM 46
2.5 PRINCIPIOS DE LA ARQUITECTURA DE LA INFORMACIÓN 48
2.6 MODELO DE MADUREZ PARA MDM 49
2.7 ESTÁNDARES PARA DATOS MAESTROS 50
2.7.1 ISO 8000 50
2.7.2 ISO/IEC 22745 56
2.8 SOLUCIONES COMERCIALES PARA LA GESTIÓN DE DATOS MAESTROS 57
2.8.1 ECCMA 57
2.8.2 PILOG 57
2.8.3 Talend MDM 59
2.8.4 Informatica MDM 60
2.8.5 IBM InfoSphere MDM 61
2.8.6 TIBCO MDM 62
2.8.7 Ataccama MDC 63
2.8.8 VisionWare Multivue MDM 65
2.9 LECTURAS RECOMENDADAS 65
2.10 SITIOS WEB RECOMENDADOS 65

CAPÍTULO 3. CALIDAD DE PROCESOS DE DATOS 67

3.1 DAMA DMBOK: DATA MANAGEMENT – BODY OF KNOWLEDGE 67
3.1.1 Áreas de conocimiento 67
3.1.2 Modelo de madurez 70
3.2 MODELO DE AIKEN 72
3.3 DATA MANAGEMENT MATURITY MODEL (DMM) 75
3.3.1 Estrategia de Gestión de Datos 77
3.3.2 Gobierno de Datos 78
3.3.3 Calidad de Datos 78
3.3.4 Operaciones de Datos 79
3.3.5 Plataforma y Arquitectura 80
3.3.6 Procesos de Soporte 81
3.3.7 Niveles de Madurez en DMM 82
3.4 MODELO DE IBM 82
3.4.1 Proceso unificado de gobierno de datos de IBM 83
3.4.2 Modelo de madurez de gobierno de datos de IBM 86
3.5 MODELO DE GARTNER DE GESTIÓN DE INFORMACIÓN EMPRESARIAL 88
3.6 TQDM 90

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:01.

3.7 DCAM 91
3.8 MODELO MAMD 94
3.8.1 Visión general del modelo 95
3.8.2 Modelo de referencia de procesos de MAMD 96
3.8.3 Modelo de evaluación 100
3.8.4 Modelo de madurez 102
3.8.5 Modelo de mejora 105
3.8.6 Comparación entre los modelos de referencia de procesos 113
3.8.7 Ejemplos de utilización de MAMD 114
3.9 LECTURAS RECOMENDADAS 124
3.10 SITIOS WEB 124

CAPÍTULO 4. CALIDAD DE REPOSITORIOS DE DATOS 125
4.1 MODELO DE CALIDAD DE DATOS 125
4.2 MEDIDAS DE CALIDAD DE DATOS 128
4.3 PROCESO DE EVALUACIÓN 131
4.3.1 Establecer los requisitos de evaluación 131
4.3.2 Especificar la evaluación 134
4.3.3 Diseñar la evaluación 135
4.3.4 Ejecutar la evaluación 137
4.3.5 Concluir la evaluación 139
4.4 CERTIFICACIÓN DE LA CALIDAD DE LOS PRODUCTOS DE DATOS 141
4.4.1 Entorno de certificación 141
4.4.2 Proceso de certificación 142
4.5 EJEMPLOS DE CERTIFICACIÓN DE CALIDAD DE DATOS 146
4.5.1 Modelo de Calidad de Datos 146
4.5.2 Proceso de Evaluación de Calidad de Datos 148
4.5.3 Certificación de Calidad de Datos 150
4.6 LECTURAS RECOMENDADAS 153
4.7 SITIOS WEB RECOMENDADOS 153

CAPÍTULO 5. MONETIZACIÓN DE LOS DATOS 155
5.1 INTRODUCCIÓN 155
5.2 CARACTERÍSTICAS ESPECIALES DE LOS DATOS 156
5.3 APROXIMACIONES A LA VALORACIÓN DE LOS DATOS 157
5.4 VALOR DE LOS DATOS 159
5.5 INFONOMÍA 162

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:01.

5.6 LECTURAS RECOMENDADAS 167
5.7 SITIOS WEB RECOMENDADOS 167

ANEXO I. PROBLEMAS DE CALIDAD DE DATOS EN BASES DE DATOS RELACIONALES 169

A1.1 PROBLEMAS A NIVEL DE ATRIBUTO / TUPLAS 169
A1.1.1 Un atributo en una tupla 169
A1.1.2 Atributos en varias tuplas (alcance de columna) 170
A1.1.3 Varios atributos en una tupla 171
A1.2 PROBLEMAS A NIVEL DE RELACIÓN 171
A1.3 PROBLEMAS A NIVEL DE VARIAS RELACIONES 172
A1.4 PROBLEMAS A NIVEL DE MÚLTIPLES FUENTES 173

ACRÓNIMOS 177

BIBLIOGRAFÍA 181

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:01.

AUTORES

ISMAEL CABALLERO MUÑOZ-REJA

Doctor Ingeniero Informático por la Universidad de Castilla-La Mancha (UCLM), Certificado CDO-1 por la Universidad de Arkansas at Little Rock y el Instituto Tecnológico de Massachusetts (MIT). CISA (Certified Information System Auditor) por la ISACA y Auditor Certificado para ISO 15504/ISO 33000 por AENOR.

Actualmente es Profesor Contratado Doctor en la Universidad de Castilla-La Mancha, acreditado de Profesor Titular de Universidad. Es uno de los socios fundadores de DQTeam SL, donde ejerce la labor de Director de Formación. Ha sido Coordinador del Máster de Big Data y Data Scientist de la Universidad Internacional de Valencia, e impartido docencia sobre gestión de calidad de datos y gobierno de datos en big data en numerosos másteres y cursos de postgrado. Ha trabajado como Secretario Técnico del Centro Mixto I+D Indra Software Labs – UCLM.

Experto Nacional por AENOR para el desarrollo de la familia de estándares ISO 8000 en el ISO TC 184/SC4/WG13, Project Leader de ISO 8000-62, ha contribuido en el desarrollo de las normas ISO 8000-60, ISO 8000-61, ISO 8000-63 e ISO 8000-66. También, ha sido experto nacional para trabajar en el ISO JTC1 WG9 de Big Data para el desarrollo de la parte de Gobierno y Calidad de Datos del estándar ISO 20547-3 sobre la Arquitectura de Referencia de Big Data.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:01.

ANA ISABEL GÓMEZ CARRETERO

Máster y Graduada en Ingeniería Informática con mención en Ingeniería del Software por la Universidad de Castilla – La Mancha. Estudiante de Doctorado en Tecnologías Informáticas Avanzadas en la Universidad de Castilla – La Mancha. Experta Universitaria en Protección de Datos por la Universidad Nacional de Educación a Distancia. CISA (Certified Information System Auditor) por la ISACA.

Ha trabajado como profesora asociada en la Universidad de Castilla – La Mancha. Actualmente trabaja como tecnóloga en el grupo Alarcos y profesora colaboradora en la Universidad Internacional de Valencia. Experta Nacional por AENOR para el desarrollo de la familia de estándares ISO 8000 en el ISO TC 184/SC4/WG13 e ISO TC 184/SC4/WG23. Socia fundadora y expresidenta de CireBits Junior Empresa y socia fundadora de DQTeam S.L.

FERNANDO GUALO CEJUDO

Máster y Graduado en Ingeniería Informática con mención en Ingeniería del Software por la Universidad de Castilla – La Mancha. CISA (Certified Information Systems Auditor) por la ISACA. Doctorando en Tecnologías Informáticas Avanzadas en la Universidad de Castilla – La Mancha. Socio fundador de DQTeam S.L. donde ejerce de director general.

JORGE MERINO GARCÍA

Doctor en Tecnologías Informáticas Avanzadas en la Universidad de Castilla – La Mancha. CISA (Certified Information Systems Auditor) por la ISACA. Ha trabajado como profesor asociado en la Universidad de Castilla – La Mancha y profesor colaborador en la Universidad Internacional de Valencia. Experto Nacional por AENOR para el desarrollo de la familia de estándares ISO/IEC 20546-7 en el ISO/IEC JTC 1/WG9.

BIBIANO RIVAS GARCÍA

Máster y Graduado en Ingeniería Informática con mención en Ingeniería del Software por la Universidad de Castilla – La Mancha. Estudiante de Doctorado en Tecnologías Informáticas Avanzadas en la Universidad de Castilla – La

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:01.

Mancha y Dottorato di Ricerca in Informatica e Matematica de la Universidad de Bari. Socio fundador de CireBits Junior Empresa y DQTeam S.L.

MARIO GERARDO PIATTINI VELTHUIS

Doctor y Licenciado en Informática por la Universidad Politécnica de Madrid. Licenciado en Psicología por la Universidad Nacional de Educación a Distancia. Máster en Auditoría Informática (CENEI), Máster en Dirección de RR.HH. (IMAFE) y Master´s Certificate en Dirección de Proyectos (George Washington University). Especialista en la Aplicación de Tecnologías de la Información en la Gestión Empresarial (CEPADE-UPM). CISA (Certified Information System Auditor), CISM (Certified Information System Manager), CRISC (Certified in Risk and Information System Control) y CGEIT (Certified in the Governance of Enterprise IT) por la ISACA. PMP (Project Management Professional) por el PMI. Diplomado en Calidad por la Asociación Española para la Calidad. Auditor Jefe ISO 15504/33000 por AENOR.

Ha trabajado como consultor para numerosos organismos y empresas, entre los que destacan: Ministerio de Industria y Energía, Ministerio de Administraciones Públicas, Siemens-Nixdorf, Unisys, Hewlett-Packard, Oracle, ICM, Atos-Ods, Avanzit, Sistemas Técnicos de Loterías, Indra/Soluziona, Alhambra/Eidos, Mundo Reader (BQ), etc. Socio fundador de las empresas Cronos Ibérica S.A (actualmente Alten), IE/Kybele Consulting S.L., AQCLab, -primer laboratorio acreditado por ENAC para la evaluación de la calidad de producto software-, y DQTeam, S.L. Ha sido profesor asociado en la Universidad Complutense y en la Universidad Carlos III de Madrid. Ha sido Director del Centro Mixto de Investigación y Desarrollo de Software UCLM-Indra, Coordinador del Área de Ciencias de la Computación y Tecnología Informática de la Agencia Nacional de Evaluación y Prospectiva (ANEP), Director del Instituto de Tecnologías y Sistemas de Información (ITSI) de la UCLM. Fundador y miembro de las Juntas Directivas de varias asociaciones científicas y profesionales, ha recibido varios premios entre los que destacan el Premio Gabriel Alonso Herrera de la JCCM a la trayectoria investigadora, el Premio Nacional a la Trayectoria Profesional del Ingeniero Informático de la Federación de Asociaciones de Ingenieros Informáticos de España, y el Premio *Aritmel* de la Sociedad Científica Informática de España como investigador, menor de 55 años, que ha realizado sobresalientes aportaciones científicas en el área de la Ingeniería Informática.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:01.

Catedrático de Universidad de Lenguajes y Sistemas Informáticos en la Escuela Superior de Informática (ESI) de la Universidad de Castilla-La Mancha (UCLM), donde dirige el grupo de investigación Alarcos, especializado en Calidad de Sistemas de Información.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:01.

PRÓLOGO

En la actualidad las organizaciones públicas y privadas han percibido la conveniencia de convertirse en "*customer centric"* (el cliente en el centro de su actividad). Ello exige que la organización se transforme digitalmente, *datificándose* y convirtiéndose en "*data centric*" y reconociendo que los datos son sus activos digitales más importantes, y por ende el centro para la toma de decisiones de las entidades y haciendo posible dar un servicio con excelencia a los *stakeholders.*

El gráfico que se incluye (Figura 0.1), resulta muy elocuente a la hora de explicar la importancia del modelo MAMD que se explica en este libro, en él se puede apreciar cómo una organización partiendo desde el concepto del *data centric* = *customer centric*, con la metodología de procesos de Gobierno, Gestión y calidad de los datos, asimismo con la evaluación de la calidad del dato de cualquier base de datos (proveniente de ERP, CRM, etc. o de SMAC+IOT+OT), logra que el dato sea adecuado para una posterior toma de decisiones: clásica, de *big data* y/o utilizando algoritmos de forma analítica (ya sea con técnicas estadísticas o de *business inteligence*-BI). Es decir, con datos de calidad e íntegros tendremos resultados de calidad para la toma de decisiones en las organizaciones en esta era digital.

En este libro se aborda, precisamente, diversos temas relacionados con este enfoque (véase figura 0.1), la calidad de los diferentes repositorios de datos (bases de datos, *big data, data lakes*, etc.) ISO 25012. Todo ello teniendo en cuenta los datos maestros y la MDM (gestión de datos maestros).

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:12.

También se aborda la parte metodológica. En este caso los diferentes procesos de gobierno, gestión y calidad de los datos-ISO 8000- y su madurez ISO 33000, que servirán para tener un Ciclo de Ingeniería del Dato *ad hoc*.

Los autores que son investigadores, profesores de universidad y profesionales del sector TIC, han desarrollado esta investigación desde hace más de 20 años, estableciendo una transferencia tecnológica de la universidad a la empresa, creando una *spinoff* (DQTeam) cuya misión y visión se ha plasmado en este libro, que supone una aportación inestimable para el sector.

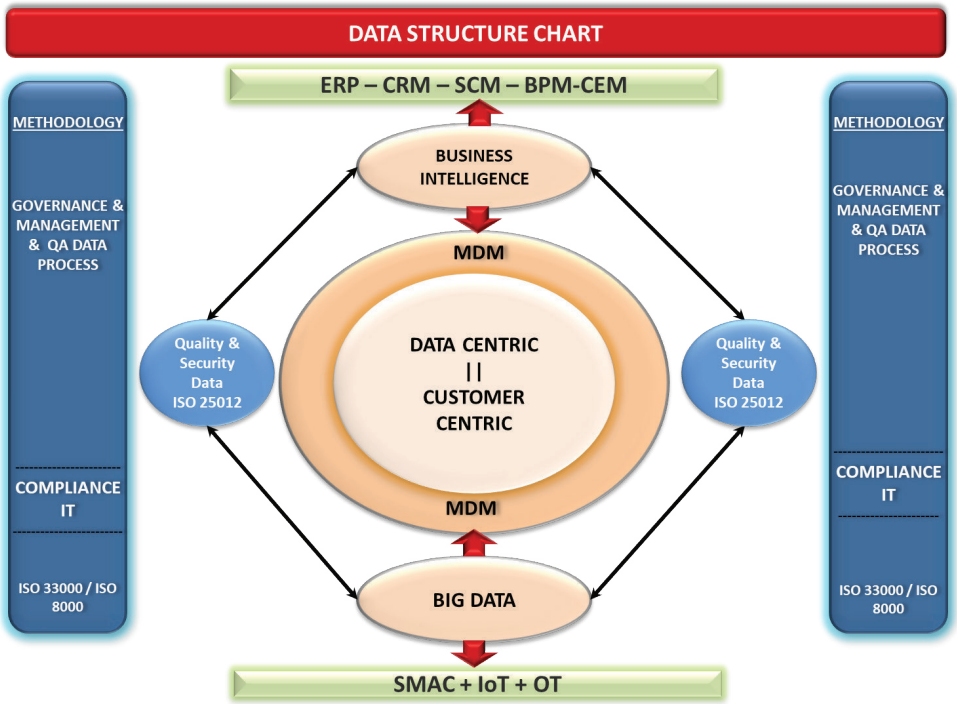

Figura 0.1. Calidad de datos y organizaciones *data/customer centric*.

Según la organización internacional IDC Research (líder mundial en proveer estudios y consultoría de inteligencia de TIC); en el año 2027, el 75% de las empresas habrán realizado la transformación digital, el futuro pues corre por estos derroteros y el presente libro bien es un manual de ayuda o guía para

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:12.

la consecución de este objetivo, pues contiene las claves y la metodología para lograrlo.

Actualmente, es de destacar que AENOR y el laboratorio acreditado de evaluación AQClab, una vez que la organización DQTeam ha realizado sus funciones de consultoría en diferentes organizaciones, ha certificado la madurez de los procesos del ciclo del dato y calidad del dato en diferentes entidades en base a unos proyectos pilotos, que se encuentran también recogidos como casos prácticos en este libro, resultando muy ilustrativo al respecto.

En sus manos tiene un libro pionero a nivel internacional, pues sirve de guía para conseguir que las organizaciones puedan lograr ***smart data*** como activo principal y cumplir con sus objetivos ya sean de negocio o de servicio público.

Sólo me queda agradecer el esfuerzo realizado por los autores y darles la enhorabuena por su investigación, síntesis y pragmatismo, pues han condensado en el libro las mejores prácticas (*best practices*) de Gobierno, Gestión y Calidad de los datos, diseñada y desarrollada por ellos en el modelo MAMD, con estándares abiertos internacionales (*open standards*-ISO).

Carlos Manuel Fernández
Asesor Estratégico de TI de AENOR
Consejero de Innovación de DQTeam

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:12.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:12.

PREFACIO

Como se señala en Ontiveros y López (2017), la primera revolución industrial se basó en la máquina de vapor, la segunda en la electricidad y el petróleo, la tercera en las tecnologías de la información y las comunicaciones; y la que estamos viviendo en la actualidad se basa en un elemento más etéreo pero omnipresente: los **datos**, que son el insumo clave de la economía del conocimiento.

Gracias a los avances que han permitido recoger y almacenar enormes cantidades de datos (ya hablamos de "zettabytes", 10^{21} bytes) por medio de diferentes tecnologías como dispositivos móviles, sensores, Internet de las cosas, *big data*, etc.; tratarlos mediante diferentes algoritmos analíticos avanzados, y disponer de una prácticamente ilimitada cantidad de procesamiento en forma de servicio, mediante la computación en la nube; los datos se han convertido en la base del funcionamiento de las organizaciones.

Este fenómeno ha coincidido también con la transformación digital de las empresas y el cambio tan profundo que está produciendo en todos los sectores económicos y en nuestra vida; transformación que tiene a los datos como su principal recurso infraestructural; y que por tanto depende fuertemente de la calidad de los datos. Todo ello ha causado también la aparición con fuerza de varios roles relacionados con los datos, el CDO (*Chief data officer*), el ingeniero/científico de datos y los DPO (*Data Protection Officer*), que se ven involucrados en diferentes aspectos del gobierno y la gestión de los datos.

La presente obra reúne diferentes aspectos relacionados con la calidad de los datos, ofreciendo una amplia visión sobre diferentes factores que se deben

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:24.

tener en consideración para la gestión, calidad y gobierno de los datos y de la información.

En este libro se persiguen los siguientes objetivos:

- Presentar de forma clara los conceptos fundamentales relacionados con la calidad de los datos.
- Dar a conocer los diferentes estándares relacionados con este tema.
- Exponer los aspectos más significativos relacionados con la calidad de los repositorios de datos.
- Analizar los diferentes procesos relacionados con la gestión, calidad y gobierno de los datos.
- Introducir algunas ideas novedosas relacionadas con la monetización de los datos.

A lo largo de esta obra se ha combinado el rigor científico con la experiencia práctica adquirida durante más de veinte años en investigación y en proyectos empresariales, proporcionando una panorámica actualizada y completa sobre la problemática asociada a la calidad de los sistemas de información.

CONTENIDO

La obra consta de cinco capítulos. El Capítulo 1 introduce los conceptos fundamentales relacionados con la calidad de datos, el gobierno y la gestión de los datos.

El Capítulo 2 resume los principales conceptos y estándares relacionados con los datos maestros, la gestión de datos maestros y su arquitectura; así como los estándares y productos más utilizados en el mercado.

En el Capítulo 3 se aborda cómo se lleva a cabo la evaluación y mejora de los procesos encargados de la gestión, calidad y gobierno de los datos. Para ello se presentan los principales modelos de referencia y madurez: DMBOK de DAMA, DMM del SEI, modelos de Aiken, IBM, EIM de Gartner, TQdM de English, DCAM del EDM, y MAMD de DQTeam.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:24.

En el Capítulo 4 se presenta un modelo junto con las características y métricas asociadas para la evaluación de la calidad de repositorios de datos (bases de datos, *big data*, *data lakes*, etc.), así como el proceso de certificación de calidad de datos.

En el Capítulo 5 se presentan algunas ideas sobre la monetización de la información, las principales aproximaciones sobre su valoración, y la infonomía.

Se incluyen también un anexo en el que se propone una clasificación de los problemas que es posible encontrar en una base de datos relacional, la bibliografía y los acrónimos utilizados en el texto.

ORIENTACIÓN A LOS LECTORES

Aunque un conocimiento en profundidad de la gestión, calidad y gobierno de los datos puede estar reservado a expertos en la materia, nuestro propósito al realizar este libro ha sido dirigirnos a una audiencia mucho más amplia que comprende:

- Alumnos de grado y postgrado de Informática, Sistemas de Información, Ingeniería del Software, Ingeniería de Servicios, Ciencia e Ingeniería de los Datos, etc.
- Participantes en seminarios o cursos monográficos sobre calidad de datos, big data o ciencia de los datos.
- Profesionales informáticos que estén trabajando en el área de datos, especialmente los CDO (*Chief Data Officers*), administradores de datos, administradores de bases de datos, directores de informática, etc.
- Directivos que tengan entre sus responsabilidades el gobierno o gestión de los datos empresariales.
- Usuarios avanzados y propietarios de procesos de negocio, que tengan interés en adquirir unos conocimientos sobre las técnicas y metodologías más utilizadas para asegurar la calidad de los datos.
- Analistas o consultores que, aun teniendo conocimientos de la materia, quieran abordarla de forma más sistemática.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:24.

Debido a la diversidad de la audiencia, el estudio de esta obra puede realizarse de maneras muy distintas dependiendo de la finalidad y conocimientos previos del lector. Se recomienda para todos los tipos de lectores comenzar por el capítulo 1, y luego ya avanzar de forma individual a los capítulos restantes dependiendo de las responsabilidades que se tengan sobre los datos.

OTRAS OBRAS RELACIONADAS

Queremos destacar que existen algunos libros que complementan la visión de la presente obra:

- *Modelo para el gobierno de las TIC basado en las normas ISO, 2ª edición. Fernández, C.M. y Piattini, M. (eds.), 2018, AENOR, Madrid.*

  Que surge como resultado de la aplicación real del modelo de AENOR de gobierno y gestión de las tecnologías y sistemas de información con estándares ISO.

- *Calidad de Sistemas de Información, 5ª edición, Piattini, M., García, F. Pino, F. y García, I., 2018, Ra-Ma, Madrid.*

  Que ofrece un tratamiento sistemático de las principales técnicas para la calidad de los procesos, proyectos, personas y servicios relacionados con los sistemas software.

- *Calidad del Software. Rodríguez, M. y Piattini, M., 2018, Ra-Ma, Madrid.*

  Que profundiza en los modelos de calidad y métricas para el producto software.

AGRADECIMIENTOS

Queremos expresar nuestro agradecimiento, en primer lugar, a todos nuestros compañeros y colaboradores de DQTeam, S.L, así como a nuestros compañeros del grupo Alarcos de la Universidad de Castilla-La Mancha. También queremos agradecer tanto a la comunidad internacional de Calidad de Datos y de Información, reunida en torno al *International Conference on Information Quality* (ICIQ), cuya edición del 2016 organizamos en Ciudad Real; como a los

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:24.

grupos de trabajo WG13 y WG23 del ISO TC184/SC4, con quien tantos hemos debatido en torno a los temas centrales del libro.

Un *ringraziamento speciale* a Domenico Natale, pionero en la estandarización de la calidad de los datos, por sus consejos y orientación.

También deseamos dar las gracias a D. Carlos Manuel Fernández Sánchez, Asesor Estratégico de AENOR y Consejero de Innovación de DQTeam por haber aceptado escribir la presentación de esta obra y por su consejo y orientación en las diferentes actividades que hemos emprendido.

Queremos también agradecer el apoyo recibido por las organizaciones y empresas en las que hemos trabajado para evaluar y/o mejorar la calidad de sus datos y los procesos de gobierno y gestión asociados; especialmente a la Biblioteca de la Universidad de Castilla-La Mancha y a su director, Antonio Galán Gall.

Por último, nos resta dar las gracias a Sandra Ramírez y Julio Santoro por sus valiosas sugerencias que, como en otras muchas ocasiones, han contribuido a mejorar considerablemente este libro, y a la editorial Ra-Ma, especialmente a Jesús Ramírez Martín y Jesús Ramírez Galán, por su apoyo y confianza.

Los autores
Ciudad Real, junio 2018

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:24.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:24.

1

FUNDAMENTOS DE LA CALIDAD DE DATOS

1.1 CALIDAD DE DATOS EN EL MUNDO DIGITAL

1.1.1 Los datos y la transformación digital

En los últimos años se viene enfatizando la importancia de los datos, y se han popularizado expresiones como: "*los datos son la nueva moneda*", "*los datos son el nuevo petróleo*", "*los datos son la mina oculta*", etc. Pero, en cualquier caso, tal como destacan Batini y Scannapieco (2016), la realidad es que "*los datos electrónicos desempeñan un rol esencial en la sociedad de las TIC*".

De hecho, la trasformación digital que está afectando a todos los sectores, desde la agricultura a la industria, el turismo, la sanidad, etc. ha convertido a los datos en el habilitador (*enabler*) más potente de cualquier tipo de organización. La razón de esto es porque los datos son la base sobre la que se toman las decisiones que afectan a las organizaciones y a las empresas. Como señala Aiken (2016), las organizaciones se ven prácticamente obligadas a estructurar en torno a los datos, las diferentes estrategias para implantar soluciones tanto tecnológicas como metodológicas para dar soporte a sus procesos de negocio. Ejemplos de dichas estrategias pueden ser: *data centricity*, *data-driven,* convergencia industrial*,* servicios híbridos y *customer centricity*. En cualquier caso, la transformación

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:37.

digital está motivando un cambio desde las estrategias "*software-centric"* a *"data-centric"*.

Todos los expertos en trasformación o disrupción digital, señalan esta importancia creciente de los datos en todos los sectores:

- Cordón et. al. (2016), dicen que "*los (big) datos son uno de los elementos más importantes de la revolución digital*".
- De la Peña y Cabezas (2015) señalan en su libro sobre transformación digital en las empresas que "*viviremos cada vez más en una avalancha de datos. Nuestro mundo se está "datificando*".
- Dobbs et al. (2015) resalta que "*la información solía ser escasa, pero la digitalización ha recortado casi a cero el coste de descubrir, transaccionar y compartir información*".
- Loucks et al. (2016) considera entre el conjunto de capacidades de la agilidad digital: la hiperactividad, la toma de decisiones informadas y la ejecución rápida. "*La hiperactividad proporciona la savia (datos y conocimientos) a los otros dos pilares. La toma de decisiones informadas, en la que los datos se analizan y distribuyen para soportar decisiones estratégicas y reglas de negocio automatizadas. que dependen de la cantidad y calidad de los datos recogidos en el estado de hiperactividad*"… destacan que "*datos malos hacen imposible una buena toma de decisiones*".
- Raskino y Waller (2015), señalan que "*la analítica y la ciencia de los datos son una de las competencias nucleares clave para la transformación digital*".
- Roger (2016), advierte que "*el papel de los datos en las organizaciones está cambiando de forma dramática en la actualidad…. Los datos están viniendo de nuevas fuentes, están siendo aplicados a nuevos problemas y convirtiéndose en impulsores (drivers) clave de la innovación*".
- Ross (2016) señala como "*la digitalización impulsó las posibilidades de la recogida de datos de una manera notable*" y que tiene gran influencia en la agricultura, sistemas financieros, etc.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:37.

- Westerman et al. (2014), destaca que "*los datos son el tipo de activo prominente en el mundo digital y requieren una atención muy significativa por parte de los ejecutivos. Los datos se han convertido en los activos digitales más valiosos para las empresas*".

- Ramírez (2017) señala como se "*requiere un cambio profundo en la mentalidad empresarial, para convertir el dato en el principal activo de la compañía y su explotación en la mayor prioridad*".

Y esta importancia se hace más patente tanto en el llamado "small" como en el "big" data ya que ambos son necesarios (como señala Ricardo Baeza en una entrevista en junio de 2017[1]), aunque es verdad que en el caso de los big data (la inteligencia de negocio y la *data science*) es todavía más elevada.

Algunos expertos incluso anuncian el surgimiento de una interesante religión: el **dataísmo**, que no venera ni dioses ni personas, sino que venera a los datos y que surge de la confluencia entre la biología y la informática (Harari, 2016).

De hecho, los avances tecnológicos que estamos viviendo en los últimos años, como pueden ser la computación en la nube, la cuantificación y monitorización de tareas cotidianas, el Internet de las cosas y la explosión de las redes sociales, han hecho que la cantidad de datos aumenten exponencialmente y se acumulen a una velocidad sin precedentes.

1.1.2 Concepto de calidad de los datos

Con los avances tecnológicos y la transformación digital, ha crecido la necesidad de usar y reutilizar los datos en los contextos más variados, aun habiendo sido definidos para un propósito específico. Esto obliga a redefinir la idea de calidad de datos desde una "adecuación al uso (*fitness for use/purpose*)" a una "adecuación a los usos (*fitness for purposes*)" (Loshin, 2013). Por tanto, está cambiando la forma de percibir la importancia y utilidad de la calidad de datos; que debe ser adaptada a diferentes contextos de uso y a diferentes sectores industriales.

1 https://insidebigdata.com/2017/07/13/big-data-small-data-correct-answer/

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:37.

También la "socialización de los datos" (*data socialization*) (Marr, 2017), destaca la importancia de asegurar que los datos adecuados está enfrente de la persona correcta (un tomador de decisiones) en el momento oportuno. Esta tendencia a "democratizar" el acceso a los datos y a que las personas interactúen colaborativamente con los mismo, también requieren un cuidado especial en su calidad.

El problema radica precisamente en cómo conseguir realmente una adecuación efectiva de los datos al contexto de uso, ya que las decisiones empresariales que pueden llegar a tomarse no serán mejores que los datos sobre los que se basan (Redman, 1996) y por tanto pueden generar ciertos errores que impactarán negativamente en la eficiencia global de la organización. Niveles inadecuados de calidad de datos supondrán: datos no usados, barreras para su accesibilidad o dificultad en su utilización en los procesos tácticos y estratégicos de posicionamiento de productos en el mercado (Strong et al. 1997). Estos inconvenientes pueden ser el origen de graves problemas organizacionales tanto desde el punto de vista nivel económico (McGilvray, 2008; Redman, 2008; Cappiello, et al. 2010), técnico (Sarsfield, 2009; Loshin, 2013), social (Eppler y Helfert, 2004), e incluso legal, baste recordar la importancia de la calidad de los datos en el nuevo Reglamento General de Protección de Datos (GPDR) de la Unión Europea, o en cualquier legislación nacional relativa a la privacidad de datos.

1.2 DATO, INFORMACIÓN, CONOCIMIENTO Y SABIDURÍA

DAMA (2017) define dato como cualquier representación de entidades o hechos que se haga mediante textos, números, gráficos, imágenes, sonido o video mientras que información, se puede entender como datos útiles en un contexto (Eppler, 2001) o valor añadido del producto de datos obtenido por su interpretación (English, 1999).

Así, tomando como base la idea de Wang (1998), se puede entender que los datos son la materia prima en un proceso de fabricación de producto de datos que llevarán a tener una información cuando se usen en un determinado contexto.

En la práctica, lo habitual es hablar indistintamente de calidad de datos o de calidad de información; aunque en sentido estricto por calidad de datos debería entenderse las características que deben tener la "materia prima"; mientras que

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:37.

por calidad de información se debería entender las características que debe tener un "producto de información" para que su utilización sea adecuada, esto es que cumpla con los requisitos y satisfaga las necesidades del usuario.

En cuanto a conocimiento es el aprendizaje que se produce al utilizar información repetidamente en un determinado contexto, con la idea de mejorar procesos productivos, reducir costes, optimizar rendimiento, etc.

La sabiduría sería el último escalón (véase figura 1.1), y se refiere a la institucionalización de un conocimiento en una organización.

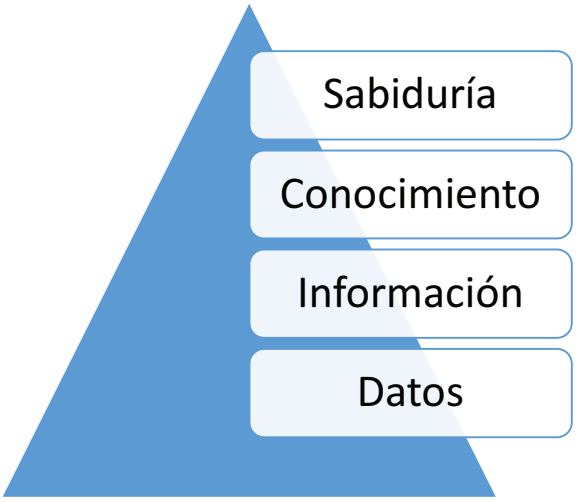

Figura 1.1. Relación entre datos, información, conocimiento y sabiduría.

Una visión complementaria y muy interesante sobre estos conceptos la proponen López y Zarza (2017), quienes consideran que *el flujo de procesos* (de datos a conocimiento) *no es unidireccional y que siempre existe un contrapunto entre nivel de detalle* (los datos con mayor nivel de detalle) *y capacidad de generalidad* (conocimiento con mayor generalidad posible). Estos expertos consideran que:

- *Los datos hacen referencia a la unidad mínima de valor respecto del análisis*, y destacan como características de los datos: origen, tipo, granularidad, estructura, incertidumbre/error y calidad.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:37.

- *La información tiene valor de referencia como propiedad agregada de los datos y sentido como producto de un proceso*. Se caracteriza por: nivel de incertidumbre, rango de validez, valor de referencia, alcance/ámbito, dominio de pertenencia, hipótesis implícitas y connotación.

- El conocimiento es la abstracción conceptual que regula el proceso instaurado para transformar los datos en información, por lo que se caracteriza por su generalización, perspectiva, eficacia y propósito.

Para López y Zarza (2017) la sabiduría no es un paso lógico del conocimiento, sino que

- Planificar (*Plan*) – Se prepara los datos que se deben utilizar.
- Obtener (*Obtain*) – Se obtienen los datos que se tienen que utilizar.
- Almacenar y Compartir (*Store and Share*) – Se mantiene una copia de los datos bien electrónicamente o bien en una copia física, y se hacen disponibles a los usuarios a través de un método de distribución.
- Mantener (*Maintain*) – Se asegura que los recursos de los datos están disponibles para el trabajo en cualquier momento.
- Aplicar (*Apply*) – Se usan los datos para alcanzar el objetivo.
- Eliminar (*Dispose*) – Se descartan los datos que ya no se van a usar

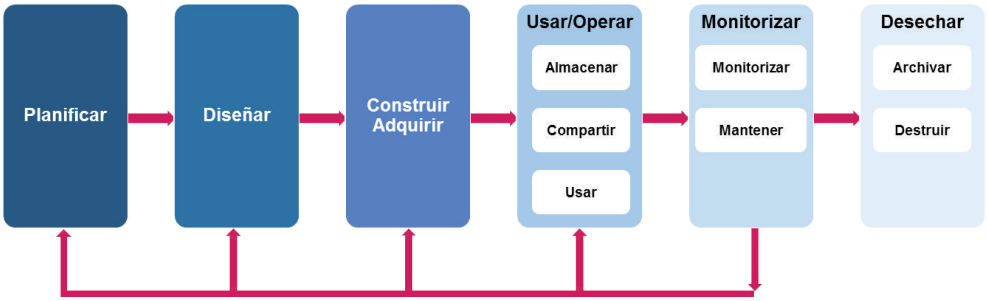

Figura 1.4. Ciclo de vida de la información en COBIT (ISACA, 2013)

Es preciso tener en cuenta que para que las organizaciones puedan utilizar los datos en sus procesos de negocio con ciertas garantías de

de capacidad para capturar adecuadamente el mundo real, bien por las propias limitaciones de los modelos y/o herramientas que se utilizan

- La organización y a su desempeño; por ejemplo, por no poner a disposición de los trabajadores los recursos de computación y formativos adecuados (Redman, 2008; English, 2009; SEI, 2013; Ladley, 2012; Infonomics, 2018).

- El diseño del proceso de negocio que utiliza los datos, entre las que se podría citar el no haber definido correctamente los requisitos de negocio necesarios para los datos (McGilvray, 2008; Cappiello et al. 2013).

- Las personas (empleados de la organización o usuarios de los sistemas de información que la organización pone a disposición de los clientes), por no tener el nivel adecuado de formación y/o experiencia para hacer una determinada tarea (Chung et al., 2002; Loshin, 2011).

- La arquitectura de los datos; como, por ejemplo, no implementar correctamente los requisitos necesarios para dar el soporte adecuado a los procesos, o tener errores en la implementación (McGilvray, 2008; Mosley et al., 2009; Tupper, 2011; West, 2011; Ladley, 2012).

- Los propios datos, que podrían no ser adecuados para el uso para el que fueron inicialmente diseñados, o dicho de modo más general, no tener características intrínsecas suficientes para ellos. (Strong et al., 1997; Loshin, 2011).

si la función de Gobierno de Datos no le proporciona los recursos necesarios y las orientaciones oportunas para alinear dichos datos con la estrategia de la organización.

Otra caracterización de problemas de calidad de datos, la proporcionan (Galway y Hanks, 2011) que en función de su naturaleza, operacional, conceptual u organizacional, recogen sus síntomas y causas (véase Tabla 1.1).

Tipo Problema	Síntomas	Causas
Operacional	Datos que no existen, no son	

- **Contexto de procesos de producción de información**. Los principales actores implicados en este contexto son los productores de información y los principales obstáculos:
  - Múltiples fuentes de datos que producen diferentes valores para el mismo atributo de la misma entidad.
  - La realización de juicios subjetivos en la producción de los datos, que puede llevar a valores diferentes.
  -

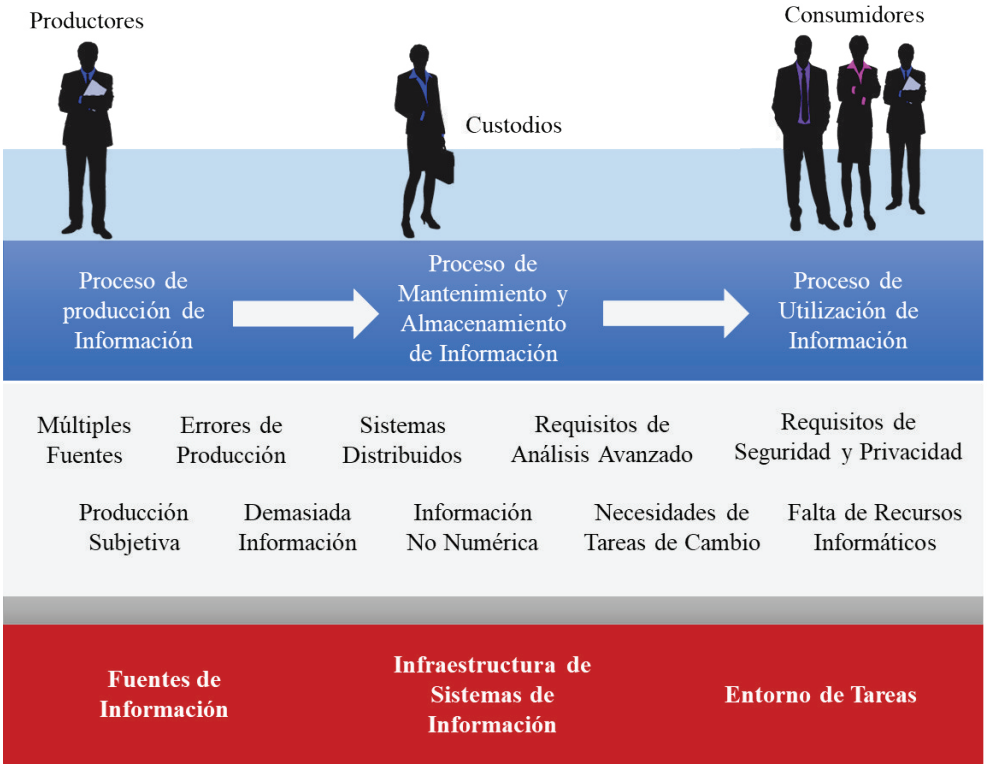

Figura 1.6. Problemas clave en calidad de datos (Strong et al., 1997)

Entrando en un nivel más tecnológico, Oliveira et al. (2005) proponen una clasificación de los problemas que es posible encontrar en una base de datos relacional, actualmente todavía las más difundidas. Este modelo (véase Anexo I) describe los problemas de calidad de datos a cuatro niveles de granularidad: varias fuentes de datos, varias relaciones (tablas), una simple relación, y nivel de atributo/tupla.

1.5 GOBIERNO VS. GESTIÓN DE DATOS

Existen multitud de definiciones de gobierno (también denominado por algunos "gobernanza") en el campo de las TICs ya que la palabra "gobierno" se ha generalizado y puesto de moda.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:37.

Desde la perspectiva de nuestra experiencia, consideramos que el objetivo general del **gobierno de datos** es maximizar el valor de los datos para la organización y gestionar los riesgos asociados, mediante la elaboración y cumplimiento de una serie de políticas de datos alineados a los objetivos

- La **Gestión de Calidad de Datos**, en base a la cultura de calidad de datos que ha sido desarrollada y manifestada a través de las políticas organizacionales de calidad de datos, tiene que proporcionar requisitos de calidad de datos a la capa de gestión de datos.

Destaca como las organizaciones con un buen gobierno de datos, ser n confiables a la hora de hacer negocios y capaces de proporcionar datos fiables, protegerán la propiedad intelectual y otros valores derivados de los datos, detectar n hackers y actividades fraudulentas, estar n m s preparadas para minimizar el impacto de las brechas de datos, ser n conscientes de cuándo y cómo pueden reutilizarse los datos, y ser n capaces de demostrar buenas pr cticas para el manejo de datos.

- ISO/IEC TR 38505-2, Governance of IT – Part 2: The implications of 38505-1 for data management (ISO/IEC, 2018), que aborda las implicaciones de la norma anterior para la gestión de los datos.

  Esta norma utiliza el marco de la anterior, con el fin de examinar las implicaciones que la ISO/IEC 38505-1 tiene para la gestión de datos, la estrategia de datos del consejo, y cómo la estrategia puede informar las políticas, los procesos y los controles relativos a los datos.

1.6 LECTURAS RECOMENDADAS

- Hidalgo, C. (2017). El triunfo de la información. La evolución del orden: de los tomos a las economías. Barcelona, Penguin Random House.

2

DATOS MAESTROS (MASTER DATA)

En este capítulo se va a aclarar los conceptos relacionados con los datos maestros y a presentar los estándares internacionales que tratan sobre los mismos, así como algunas soluciones existentes en el mercado.

2.1 INTRODUCCIÓN A LOS DATOS MAESTROS

Los Datos Maestros hacen referencia a aquellos objetos críticos en las organizaciones que representan aspectos clave del negocio como pueden ser clientes, productos, empleados, proveedores, etc. Estos objetos están formados por un conjunto de datos que representan una entidad específica desde distintos puntos de vista de la operativa del negocio. Por ejemplo, para el objeto cliente, existirán una serie de aspectos muy relevantes para según qué perspectiva del negocio, así en el departamento de envíos existirá una base de datos que almacene datos precisos sobre los datos de contacto del cliente lo más precisa posible, mientras que en el departamento de ventas existirá una base de datos sobre el contacto del cliente (véase Figura 2.1).

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:51.

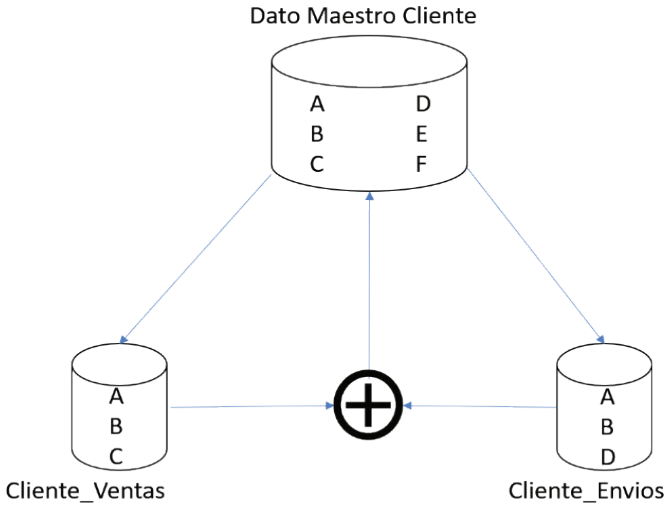

Figura 2.1. Datos Maestros

Algunos autores utilizan los términos datos de referencia y datos maestros de forma intercambiable. Pero no todos los datos de referencia son datos maestros. Se puede definir dato de referencia como cualquier dato que se puede utilizar para validar, clasificar o categorizar otros datos, por ejemplo, el código de país según ISO, o códigos postales. Los datos de referencia se usan para validar datos de entrada o como lista de valores correctos.

Una de las ventajas más importantes de los datos maestros es asegurar que los procesos de la organización no fallan debido a la inconsistencia en la definición de los datos entre las distintas bases de datos de las distintas operativas del negocio o porque no exista una consistencia al agregarse datos en los sistemas analíticos de la organización o al consolidarse en almacenes de datos. Por ejemplo, supóngase que un cliente cambia de dirección postal y este cambio se actualiza sólo en la base de datos Cliente_Ventas. Posteriormente si la organización realiza un envío al cliente, pero el sistema que realiza el envío opera con la base de datos Cliente_Envios (diferente de la de Cliente_Ventas), el producto se enviará a la dirección equivocada, y esto supondrá varios perjuicios importantes: se pierde el valor de lo enviado, se pierde la comunicación con el cliente, y se puede perder la confianza del cliente.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:51.

Los datos maestros se pueden identificar y modelar mediante un enfoque descendente, analizando qué conceptos se comparten entre los diferentes procesos de negocio; o mediante un enfoque ascendente evaluando los activos de datos desde un bajo nivel, identificandolas aplicaciones que usan datos candidatos a ser instancias de lo que serán datos maestros.

En definitiva, permiten a la organización crear y usar una "única versión de la verdad" y resolver los problemas de calidad de datos de una manera más integral a lo largo de toda la organización (Berson y Dubov, 2011). Algunos ejemplos de datos maestros ya utilizados en las empresas son los datos de clientes se denomina CDI (*Customer Data Integration*) y cuando se centra en los productos (*Product Information Master*).

Precisamente en el "*The Business Agility Manifesto. Building for Change*"[2], Roger Burlton, Ronald Ross y John Zachman plantean la "Directiva Suprema" como "*Todas las iniciativas deben estar alineadas demostrablemente con los Imperativos de la Gestión*", y dentro del desarrollo de esta directiva tratan específicamente de la "fuente única de la verdad de la organización":

2.2 GESTIÓN DE DATOS MAESTROS

La gestión de datos maestros, conocida por sus siglas en inglés MDM (*Master Data Management*), es un conjunto de buenas prácticas por parte de los principales *stakeholders* (la organización, usuarios del sistema, …) y aplicaciones, soportado por distintas herramientas y metodologías que permiten establecer las políticas, procedimientos, servicios e infraestructura para lograr la captura, integración y uso completo de los datos maestros, véase Figura 2.2.

Copyright © 2018. RA-MA Editorial. All rights reserved.

2 https://busagilitymanifesto.org/?utm_source=brs_top1&utm_medium=email&utm_campaign=BRS057_20180122

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:51.

Figura 2.2. Elementos de un programa de MDM (Loshin, 2010)

Berson y Dubov (2011) definen MDM como "*el marco de los procesos y tecnologías que tienen como objetivo crear y mantener un entorno de datos acreditado, fiable, sostenible, preciso y seguro que represente un "*única y holística versión de la verdad" *para los datos maestros y sus relaciones, así como un banco de pruebas (benchmark) aceptado y utilizado dentro de la organización y entre organizaciones y que comprenda un conjunto diverso de sistemas de aplicación, líneas de negocio, canales y comunidades de usuario*".

Gartner[3], por su parte, define la gestión de datos maestros como *"una disciplina tecnológica en la que el negocio y las tecnologías de la información trabajan juntos para asegurar la uniformidad, precisión, administración, coherencia semántica y responsabilidad de los activos de datos maestros compartidos oficiales de la empresa*".

Mediante un programa de gestión de datos maestros se ofrece acceso a vistas consistentes de las entidades de datos maestros identificables de forma única en la infraestructura operacional de la aplicación, más concretamente se logran los siguientes objetivos (Loshin, 2008):

- Creación de un único repositorio de datos maestros de alta calidad.
- Alimentación de aplicaciones a través de la organización con una visión sincronizada y consistente de los datos de la empresa.
- Integración de alta calidad de las instancias de datos maestros en la empresa.

3 http://www.gartner.com/it-glossary/master-data-management-mdm

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:51.

Entre los beneficios que lleva consigo la MDM, cabe destacar:

- Facilitación del gobierno del dato.
- Mejora de la calidad de los datos.
- Conocimiento en profundidad del cliente.
- Mejora del servicio al cliente.
- Presentación consistente de informes.
- Mejora de la competitividad, la gestión de riesgos, la eficiencia operativa, reducción de costes, productividad del negocio y la toma de decisiones.
- Mejor planificación y análisis de gastos.
- Mejor planificar y previsión de futuros cambios (sabiendo qué

real en sus representaciones en las diversas bases de datos (Talburt, 2011). Este proceso consta de los siguientes pasos:

1. Reconocer las fuentes de los datos maestros.
2.

permanecen distribuidos en varios sistemas. Este enfoque es el único que ofrece una integración puramente virtual.

- La última variante (v4) es un sistema P2P (*Peer-to-Peer*), basado en un enfoque punto a punto. Aquí el sistema de información

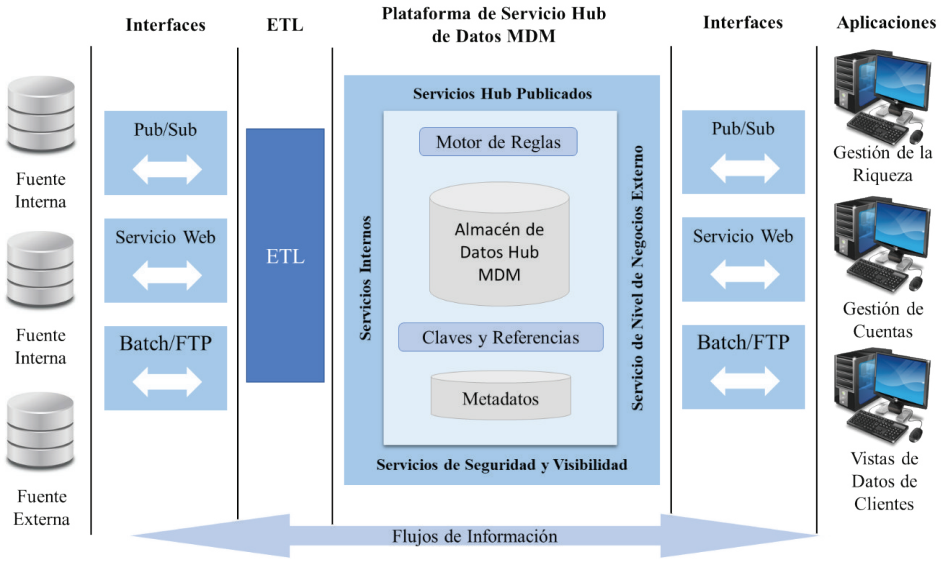

Figura 2.4. Arquitectura detalla para MDM

2.5 PRINCIPIOS DE LA ARQUITECTURA DE LA INFORMACIÓN

Gartner identifica varios principios para crear una arquitectura de la información empresarial.

- P3. Todas las modificaciones y correcciones a los datos maestros pueden realizarse solo de acuerdo a las reglas y políticas establecidas por la organización, incluyéndolas reglas para resolver conflictos de cambios de datos, estos cambios

- **Nivel inicial**

  El nivel inicial de madurez existe pocas posibilidades para explotar los datos maestros.

- **Nivel reactivo**

  En este nivel se intenta resolver el problema que representa la existencia de copias de datos, se considera que los datos duplicados son un problema tecnológico y se compran algunas herramientas de calidad de datos.

- **Nivel gestionado**

  Se crean algunas aplicaciones analíticas que se basan en un cierto nivel de consolidación. La capacidad de utilizar datos maestros se convierte en un proceso repetible que se utiliza para incorporar nuevas aplicaciones. Se dispone de servicios de consolidación y sincronización.

- **Nivel proactivo**

  Las aplicaciones se integra

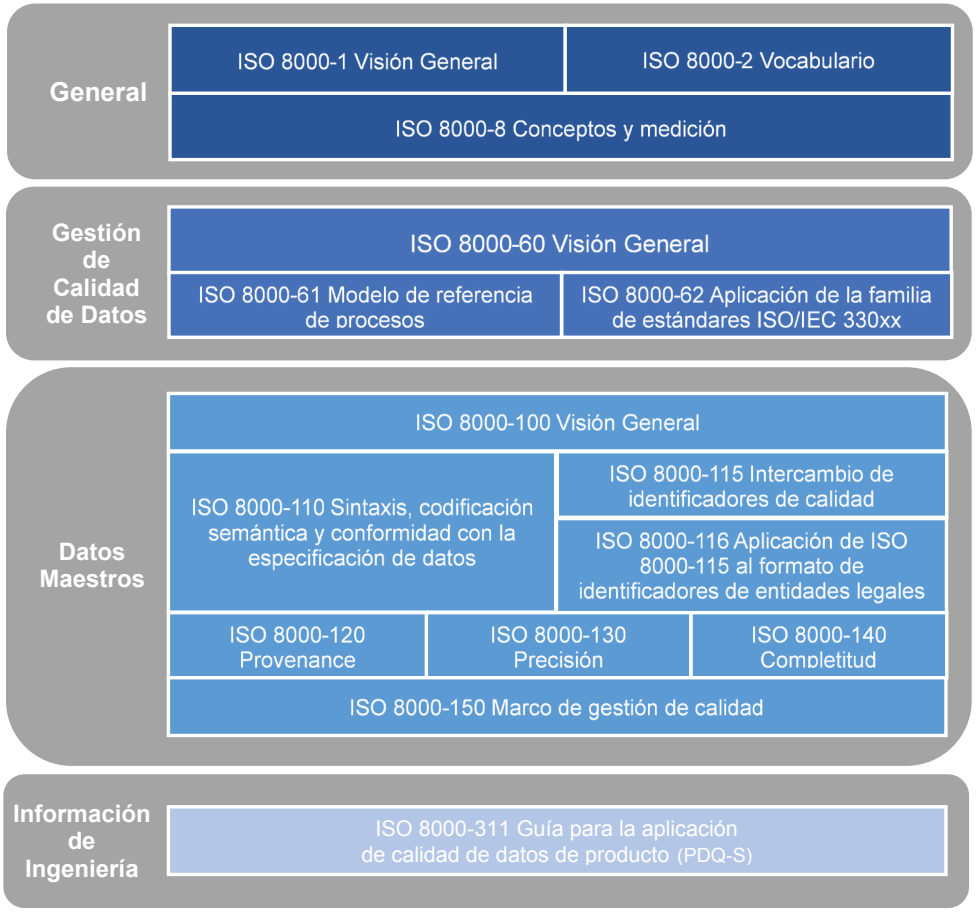

Figura 2.5. Visión General ISO 8000

- **General.**

la madurez organizacional de los procesos de gestión de calidad de datos.

- **Información de ingeniería.** La parte de información de ingeniería sólo contiene la norma ISO 8000-311. Es una guía para la aplicación de la calidad de los datos del producto para (PDQ-S), como se describe en ISO 10303-59.

- **Datos maestros.** La serie ISO 8000-1x0 aborda principalmente el intercambio de datos maestros, por lo que es en la que profundizaremos en este capítulo.

Las partes 100 a 150 de la norma describen algunos aspectos sobre el intercambio de datos maestros entre aplicaciones (u organizaciones) por medio de mensajes de datos maestros. Para ello proponen usar un formato preestablecido para construir los mensajes y para encapsular los datos maestros dentro de los mensajes. Además, se especifican los requisitos que deben cumplir dichos mensajes para incorporar información sobre el nivel de precisión y completitud de los datos contenidos en el mensaje de datos maestros.

Los roles que intervienen en un escenario de intercambio de datos maestros según el estándar son:

- **Proveedores de datos:** entienden las necesidades de datos del resto de los usuarios y los requisitos de calidad sobre ellos que se espera de ellos. Por ejemplo, una agencia de viajes.

- **Consumidores (clientes de datos):** demandan datos con un cierto nivel de calidad. Por ejemplo, un cliente web.

- **Custodios de datos:** se encargan entre otras tareas de gestión, del almacenamiento de los datos, teniendo en cuenta los requisitos de acceso seguro a los datos, así como los requisitos necesarios para servir datos con la calidad demandada por los clientes.

ISO 8000-100

Esta parte de la familia ISO 8000 contiene una visión global de los estándares que abordan la cuestión de los datos maestros, incluyendo:

- Especificación del alcance de los datos maestros.
- Introducción al concepto de datos maestros.
- Descripción de la arquitectura de datos.
- Visión global de los contenidos de otras partes del estándar.

En la Figura 2.6 se muestra la arquitectura de los Datos Maestros propuesta en la ISO 8000-100.

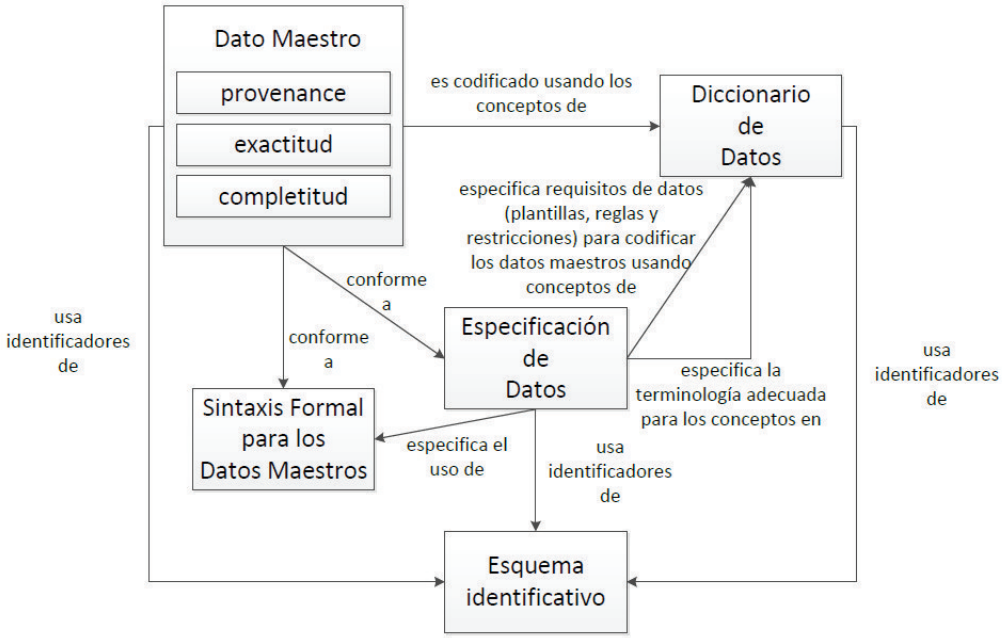

Figura 2.6. Arquitectura de los Datos Maestros (tomado de ISO 8000-100)

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:51.

ISO 8000-102

Esta parte contiene el vocabulario relacionado con la gestión del intercambio de datos maestros referente a las distintas partes de la norma ISO 80001x0; se definen entre otros conceptos los siguientes (ISO, 2009a):

- Conjunto de datos (*data set*): agrupación significativa de datos.
- Datos maestros (*master data*): datos que posee una organización que describe las entidades que son independientes y fundamentales para esa organización, y que es necesario referirse a ellos para realizar sus transacciones.
- Diccionario de datos (*data dictionary*): colección

- Registro de procedencia del dato (*data provenance record*): registro en la última descendencia y transcurso de una parte de un dato a

ISO 8000-130

Esta parte (ISO, 2016c) es una adición opcional a la parte 120, que describe cómo añadir información sobre el nivel de precisión de los datos al mensaje de datos maestros.

ISO 8000-140

Análogamente a la anterior, esta parte (ISO, 2016d) especifica requisitos para la representación y el intercambio de información acerca de la completitud de los datos maestros, incluyendo los requisitos de captura e intercambio de información de la compleción de los datos en forma de representación/declaración y el aseguramiento de la compleción de los datos, adem s de un modelo de datos conceptual para la información de compleción de los datos.

ISO 8000-150

Esta parte (ISO, 2011), incluye un marco de trabajo que identifica los procesos para la gestión de la calidad de los datos maestros. Este marco se puede usar junto con, o independientemente de, estándares de sistemas de gestión de calidad como, por ejemplo, ISO 9001.

2.7.2 ISO/IEC 22745

ISO/IEC 22745 es un est ndar

- Parte 13: Identificación de conceptos y terminología.
- Parte 14: Interfaz de consultas al diccionario.
- Parte 20: Procedimientos para el mantenimiento de un diccionario t cnico abierto.
- Parte 30: Identificación de guías para la representación.
- Parte 35: Consultas para datos maestros.
- Parte 40: Representación de los datos maestros.

2.8 SOLUCIONES COMERCIALES PARA LA GESTIÓN DE DATOS MAESTROS

2.8.1 ECCMA

La ECCMA (*Electronic Commerce Code Management Association)* fue fundada en 1999 con el objetivo de desarrollar y mantener de forma colaborativa un diccionario est ndar de manera abierta y r pida. La idea es mantener un vocabulario coherente y operativo con los t rminos necesarios para que las organizaciones que se suscriben a sus servicios puedan intercambiar datos sin perder su significado.

Para ello, instan a los gestores de programas de gobiernos de datos que asimilen los principios de Diccionario de Datos establecidos en los est ndares ISO 8000-110 y a usar el diccionario de datos formateado de acuerdo a los principios de ISO 22745.

ECCMA proporciona una serie de servicios Web para la creación y consulta de los datos maestros, aunque, no proporciona servicio de medición y etiquetado de los niveles de calidad de datos de los datos maestros.

2.8.2 PILOG

PiLog es un proveedor global de soluciones de calidad de datos maestros basados en est ndares internacionales, para poblar sistemas ERP, que provee servicios de gestión de datos maestros, incluyendo limpieza de datos y de

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:51.

clasificación y mantenimiento a través de sus herramientas y aplicaciones. Estas herramientas (Figura 2.7) se han diseñado conforme a la norma ISO 8000, el est ndar internacional de Calidad de Datos Maestros.

Entre sus productos destacan los siguientes:

- **Diccionario Técnico**: El Diccionario T cnico de PiLog (OTD) proporciona un conjunto sólido de reglas básicas para cada clase de datos, y est precargado con miles de plantillas de estructuración de datos ORM listas para usar.

- **Master Data Record Manager (MDRM)**: Es una herramienta de gestión de datos maestros con flujos de trabajo para guiar el proceso de agregar, cambiar, aprobar y eliminar registros correspondientes a datos maestros.

- **Master Data Acquisition Manager (MDAM)**: est desarrollado para adquirir, construir, estructurar, limpiar y configurar los datos. Las tareas de mantenimiento se pueden crear y vincular al objeto de mantenimiento pertinente para garantizar los procedimientos de trabajo de mantenimiento de la implementación, la planificación, la presupuestación, la planificación de recursos, los requisitos de inventario y la disponibilidad optimizada de los equipos operativos.

- **Contenido de los Productos (PiLog Preferred Ontology)**: PiLog ofrece una ontología integral "lista para usar" conocida como PiLog Preferred Ontology (PPO). El PPO es un conjunto de requisitos de datos (plantillas), propiedades asociadas, unidades de medida, tipos de datos y reglas (m scaras, expresiones regulares) que se han probado en el mercado durante casi 15 años. PiLog proporciona a la ontología un servicio de software de gestión que permite a los clientes ensamblar r pidamente su propio conjunto de ontologías preferidas corporativas (CPO) del PPO, el cliente puede adoptar el PPO tal cual o realizar cambios para satisfacer sus necesidades específicas.

- **Master Data Ontology Manager (MDOM):** permite crear y mantener el diccionario corporativo de la organización, los requisitos de datos y clasificación de estructuras. Los requisitos de datos están integrados con el diccionario que contiene los términos y las definiciones, así como sinónimos y traducciones.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:51.

- **PiLog Preferred Record Manager (PPRM):** PiLog ha reunido bibliotecas de registros de datos maestros estructurados y conjuntos de valores de datos en dominios de datos maestros específicos tales como maestros de materiales, que pueden usarse para aumentar los conjuntos de registros de datos maestros de clientes existentes o como un inicio r pido para crear una base desde donde trabajar. Los conjuntos de "registros preferidos" de Pilog incluyen conjuntos de valores est ndar, asignaciones entre conjuntos de valores estructurados y no estructurados y listas controladas específicas para la propiedad de datos de un requisito de datos (por ejemplo, lista de códigos postales).

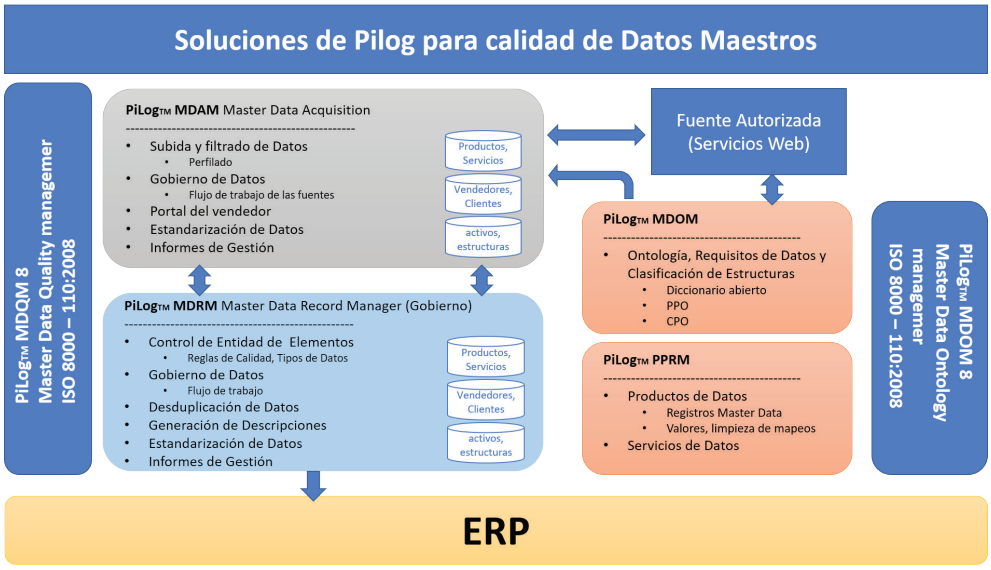

Figura 2.7. Entorno Master Data Quality Solutions de Pilog

2.8.3 Talend MDM

Talend MDM es una herramienta de desarrollo y diseño basada en Eclipse que permite operaciones *drag-and-drop* para el enrutado de datos y creación de servicios web. Est basado en un modelo de datos activo basado en XML que permite modelar cualquier dominio de datos sin estar restringido a estructuras predefinidas. Tiene un modelo de precios de subscripción, así como una versión gratuita m s limitada con *GPL open source license*.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:51.

Ofrece una solución multidominio, integración de datos (incluyendo grandes volúmenes de datos y servicios basados en la nube), así como capacidades de calidad de datos (perfilado de datos, automatización de resolución de errores de calidad de datos y aplicación de reglas). Talend favorece el modelado visual y la importación /exportación de modelos de datos, así como flujos de trabajo integrados para la administración y gobierno de datos; lenguaje de consulta MDM con acceso a datos REST y repositorio de datos maestros.

Ofrece m s de 400 conectores con diferentes tipos de fuente de datos o aplicación consumidora como: Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform, Oracle, Teradata, Microsoft SQL Server, Marketo, Salesforce, NetSuite, así como herramientas ETL de alto rendimiento que facilitan la carga de datos por lotes o tiempo real para importar, exportar y combinar datos de varios tipos de base de datos (Excel, CSV, Parquet, AVRO, HDFS) y herramientas de limpieza de datos para eliminar registros duplicados y enrutamiento autom tico.

Tambi n proporciona otras capacidades avanzadas de integración de datos que incluyen la captura de datos, integración de reglas JBoss y un administrador SOA que facilita la exposición de operaciones de datos maestros como servicios web y mecanismos de colaboración empresarial que incluyen LDAP, Active Directory y la integración de SSO; privilegios de acceso basados en roles; y herramientas de automatización del flujo de trabajo para ayudar a diseñar e implementar procesos de gobernanza multi-role para la gestión de datos maestros.

Proporciona tambi n capacidades de despliegue de ESB (*Enterprise Service Bus*) basado en OSGI y Spring Boot como microservicios y su gestión mediante monitorización JMX y estadísticas de la actividad de flujo de los mensajes.

2.8.4 Informatica MDM

Informatica MDM ofrece una solución de gestión de datos maestros modular que agiliza los procesos de centralizar y mantener la gestión de datos de clientes durante todo el ciclo de vida, permitiendo conectar a los clientes con nuevos tipos de datos o relaciones y mantenerlos actualizados.

Informatica MDM permite adaptarse a requisitos específicos ofreciendo:

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:51.

- Visión única de los datos: la creación de una visión única y fidedigna de los datos críticos para el negocio a partir de información dispar, duplicada y contradictoria.
- Visión integral de las relaciones: las reglas de negocio permiten identificar las relaciones entre los datos.
- Visión completa de todas las interacciones: la integración de las transacciones e interacciones de redes sociales que han tenido lugar con ese producto, cliente, socio de canal o cualquier otro elemento de datos le proporciona una visión completa de ese cliente.

2.8.5 IBM InfoSphere MDM

Es una solución MDM completa y flexible para mejorar los procesos operativos de negocio, compatible con la mayor parte de los dominios, estilos arquitectónicos y casos de uso en todas las industrias. La solución est optimizada para grandes volúmenes de datos e integración en tiempo real y proporciona la capacidad de comprender mejor los comportamientos de los clientes, permitiendo una mejor experiencia del cliente.

IBM InfoSphere MDM ofrece una Integración mejorada con IBM DataStage para extracción de datos y f cil publicación de datos MDM en sistemas heredados y procesos comerciales, destacando:

- Motor de coincidencia probabilística, administración de duplicados y la visualización avanzada de relaciones y jerarquías orientadas a tareas.
- Potenciar los programas de negocio impulsados por datos y la colaboración en uno o varios equipos de una manera gobernada y conforme.
- Servicios multidominio y modelos de datos que permiten un uso m s amplio de los datos, unificando fuentes dispares de datos e integración en múltiples repositorios.
- Flexibilidad de implementación.
- Rendimiento y escalabilidad.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:51.

- La gestión de la estrategia de la información, la estrategia de producto y funcionalidad.
- Servicio de integración.
- Enfoque industrial (sector público de la salud).

2.8.6 TIBCO MDM

Es una solución de gestión de datos maestros multidominio, impulsada por eventos, diseñada específicamente para entornos de distribución de múltiples niveles. TIBCO MDM proporciona una plataforma para crear y mantener una versión única, completa y confiable de datos sobre sus clientes, proveedores, activos físicos, organización, socios, productos, materiales, empleados y otros recursos.

Al consolidar, quitar la duplicación y limpiar la información ubicada en múltiples silos de aplicaciones diseminados por toda la empresa, TIBCO MDM permite crear una fuente confiable de datos maestros críticos y acceder en tiempo real para tomar decisiones comerciales m s informadas. TIBCO MDM est pensado para una eficaz gestión de múltiples dominios de datos, destaca por su solución basada en la nube y por Visual MDM, una característica diferenciadora en cuadros de mando de calidad de datos. Otras fortalezas son:

- Proporcionar una extensa lista de conectores para la ingesta de datos de fuentes de datos sociales, streaming e IoT.
- Automatizar y simplificar la coincidencia de datos y la deduplicación con un motor integrado de aprendizaje autom tico.
- Una API y un entorno de desarrollo f ciles de utilizar que permite a cualquier persona participar en el proceso de MDM a trav s de cualquier aplicación comercial, portal web o aplicación móvil.
- Un repositorio de datos maestros ampliable con características flexibles que proveen una vista centralizada de todas las relaciones entre tipos de datos, aclarando complejas relaciones entre dominios.
- Sistema publicador/subscritor seguro.
- Plantillas listas para usar.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:51.

- Soporte a cuadros de mandos.
- Altamente escalable.
- Interfaz *drag-and-drop* que permiten diseñar interfaces de usuario m s simples y limpias.

2.8.7 Ataccama MDC

Es una plataforma avanzada, escalable y altamente disponible para consolidar y proporcionar tareas de administración de datos maestros. Es un modelo y un sistema de procesamiento de datos impulsado por metadatos que mejora y complementa el motor de calidad de datos de Ataccama. Destaca por su Data Center (MDC), e instalaciones de alta disponibilidad, se centra en las aplicaciones de negocio de MDM, la comprensión de la industria y el apoyo a la gobernabilidad de datos maestros. Su precio es inferior a la media.

Adem s del motor de *back-end* de MDC, Ataccama Master Data Admin (MDA) permite a los usuarios realizar todas las tareas críticas de administración de datos, no solo navegar, modificar, buscar y crear datos maestros a través de una interfaz web, sino también a navegar y administrar jerarquías gráficas.

Las características clave incluyen:

- Aseguramiento de la calidad de los datos: todas las entradas del administrador están sujetas a la validación del motor DQ, que actúa como un firewall DQ (Data Quality) y evita la creación de discrepancias y duplicidades.
- Visualizaciones: Proporciona toda la información necesaria para ayudar a los administradores de datos en las tareas cotidianas. Incluye una interfaz para monitorización de calidad de datos e informes de tendencias (DQ Dashboard), problemas de calidad de datos y excepciones resolución (Rastreador de problemas de DQ.
- Gestión de problemas: Proporciona a los administradores de datos una lista de tareas y problemas relacionados con los registros maestros que experimentan errores, y sugiere soluciones, como propuestas de coincidencia o valores correctos.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:51.

- Soporte para múltiples escenarios.
- Integración de datos del cliente (CDI), Gestión de prospectos, Consolidación del catálogo de productos, Gestión de información del producto (PIM).
- Administración de datos y análisis.
- Una interfaz de usuario f cil de usar para usuarios empresariales con varios diseños de datos intuitivos y visualizaciones, entrada de usuario controlada, flujos de trabajo de aprobación simples y gestión de problemas.
- Soporte para todos los estilos de arquitectura comunes.
- Estilo de consolidación, estilo de coexistencia, estilo centralizado (sistema de registro), mezcla de estilos.
- Alta disponibilidad, coherencia transaccional, monitoreo, auditoría, seguridad y permisos.
- Basado en modelos y metadatos.
- Administración de front-end: Los administradores de datos tienen la capacidad de guardar registros importantes o interesantes para futuras acciones y acceder f cilmente a los borradores.
- Integración, SOA, mensajería: La integración con otros sistemas fuente y de consumo a trav s de interfaces por lotes (archivos, JDBC), interfaces en línea (HTTP, JMS) y mensajería (JMS) es f cil de configurar. MDC puede leer y escribir archivos Hadoop e integrarse con las interfaces Stream y REST.
- MDC admite la transformación de la calidad de los datos a la gestión de datos maestros. Cualquier solución de calidad de datos o gestión de datos creada con las herramientas de Ataccama se puede usar r pida y f cilmente como base de su MDM hub.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:51.

2.8.8 VisionWare Multivue MDM

Es una buena opción para organizaciones que usan software de Microsoft, que ofrece mejoras en el núcleo de MDM. Utiliza MongoDB como almac n de datos, con sincronización opcional de su capa RDMBS, y soporta el despliegue de las plataformas en la nube como Amazon Web Services y Microsoft Azure. Entre sus características destacan:

- Sofisticados algoritmos de coincidencia.
- Herramientas robustas de gobierno de datos.
- Herramientas potentes para usuarios de negocios y la capacidad de integrarse con sistemas CRM.
- Proceso de implementación predecible y altos niveles de escalabilidad.
- Posibilidad de implementar en las instalaciones o en la nube.

2.9 LECTURAS RECOMENDADAS

- D. Loshin. (2010). *Master Data Management.* Morgan Kaufmann.

  En este libro, se ilustra las claves de xito en MDM, por lo que ayuda a tener en cuenta todos los detalles involucrados en la planificación y ejecución de un proyecto MDM que produzca mejoras en la productividad y efectividad de la organización.

2.10 SITIOS WEB RECOMENDADOS

- *www.eccma.org*

  Web de la *Electronic Commerce Code Management Association*

- *www.piloggroup.com/master-data-governance/*

  Web de PiLog Group

- *es.talend.com*

  Web de Talend

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:51.

- *www.informatica.com/mx/products/master-data-management/multidomain-mdm.html*

  Web de Informatica

- *www.ibm.com/us-en/marketplace/ibm-infosphere-master-data-management*

  Web de IBM MDM

- *www.tibco.com/resources/whitepaper/tibco-master-data-management-platform*

  Web de TIBCO MDM

- *www.ataccama.com/products/master-data-management/mdc*

  Web de Ataccama MDC

- *www.visionware.com/mdm/*

  Web de VisionWare Multivue MDM

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:42:51.

3

CALIDAD DE PROCESOS DE DATOS

Al igual que sucede en el campo del software, también para los datos se han definido modelos de evaluación y mejora de la madurez/capacidad para los procesos que abordan las disciplinas de gestión de datos, gestión de calidad de datos y gobiernos de datos. En este capítulo se resumen modelos de referencia de procesos, así como los modelos de madurez más relevantes.

3.1 DAMA DMBOK: DATA MANAGEMENT – BODY OF KNOWLEDGE

3.1.1 Áreas de conocimiento

La *Data Management Association* (DAMA) propone un modelo de referencia que agrupa los procesos en áreas de conocimiento de gestión de datos (DMBOK, *Data Management BOdy of Knowledge*) (DAMA, 2017) que se representan en una rueda en la que sitúa el gobierno de datos en el centro (Figura 3.1); debido a que en el modelo se entiende el gobierno de datos como el área que da consistencia a y equilibro entre todas las funciones. Estas áreas de conocimiento son:

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

Figura 3.1. Framework de gestión de datos de DAMA (DAMA, 2017)

- **Gobierno de datos**, proporciona la dirección y supervisión para la gestión de los datos, estableciendo un sistema de derechos de decisión sobre los datos que responde a las necesidades de la organización.
- **Arquitectura de datos,** define los diseños arquitectónicos para la gestión de los activos de datos alineándolos con la estrategia organizacional, para establecer los requisitos estratégicos de datos y los diseños que satisfagan dichos requisitos.
- **Diseño y modelado de datos**, es el proceso de descubrimiento, análisis, representación y comunicación de los requisitos de datos de una forma precisa ("modelo de datos").
- **Almacenamiento y operaciones de datos,** incluye el diseño, implementación y soporte de los datos almacenados para maximizar su valor a lo largo de su ciclo de vida, desde la planificación hasta la retirada de los datos.
- **Seguridad de datos,** incluye la planificación, desarrollo y ejecución de las políticas y procedimientos de seguridad para proporcionar autenticación, autorización, acceso y auditoría de los activos de datos.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

- **Integración e interoperabilidad de datos,** incluye los procesos relacionados con la transferencia y consolidación de los datos entre almacenes de datos, aplicaciones y organizaciones. La integración consolida los datos de forma consistente tanto física como virtualmente, mientras que la interoperabilidad es la habilidad de comunicar los datos entre múltiples sistemas.

- **Gestión de documentación y contenido**, incluye la planificación, implementación y actividades de control para la gestión del ciclo de vida de los datos y la información que se encuentran en cualquier forma y en cualquier medio; especialmente documentos necesarios para soportar los requisitos de cumplimiento legal y normativo.

- **Datos de referencia y datos maestros**, incluye la reconciliación y mantenimiento continuos de los datos clave y críticos compartidos para permitir el uso consistente entre los diferentes sistemas de la "versión de la verdad" de las entidades esenciales de la organización, más exacta, oportuna y relevante.

- **Almacenes de datos e inteligencia de negocio**, incluye los procesos de planificación, implementación y control para gestionar los datos de soporte a las decisiones y habilitar a los "trabajadores del conocimiento" para obtener valor a partir de los datos por medio del análisis y reporte.

- **Metadatos,** incluye las actividades de planificación, implementación y control para permitir el acceso a metadatos integrados de alta calidad, incluyendo definiciones, modelos, flujos de datos y otra información crítica para entender los datos y los sistemas por los cuales se crean, mantienen y acceden.

- **Calidad de datos**, se refiere a la planificación, e implementación de técnicas de gestión de calidad para medir, acceder y mejorar la adecuación de los datos para su uso en la organización.

En DMBOK para cada una de estas áreas se presentan sus actividades, herramientas, técnicas, guías de implementación y métricas, así como bibliografía recomendada. Además, se incluyen capítulos sobre la ética en la manipulación de datos, *big data* y la ciencia de los datos, la evaluación de la madurez de la gestión de los datos, la organización y los roles para la gestión de los datos, y la gestión del cambio organizacional y la gestión de los datos,

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

3.1.2 Modelo de madurez

En cuanto a la evaluación de la madurez de la gestión de los datos, DAMA propone el siguiente esquema de proceso representado en la Figura 3.2 (DAMA, 2017).

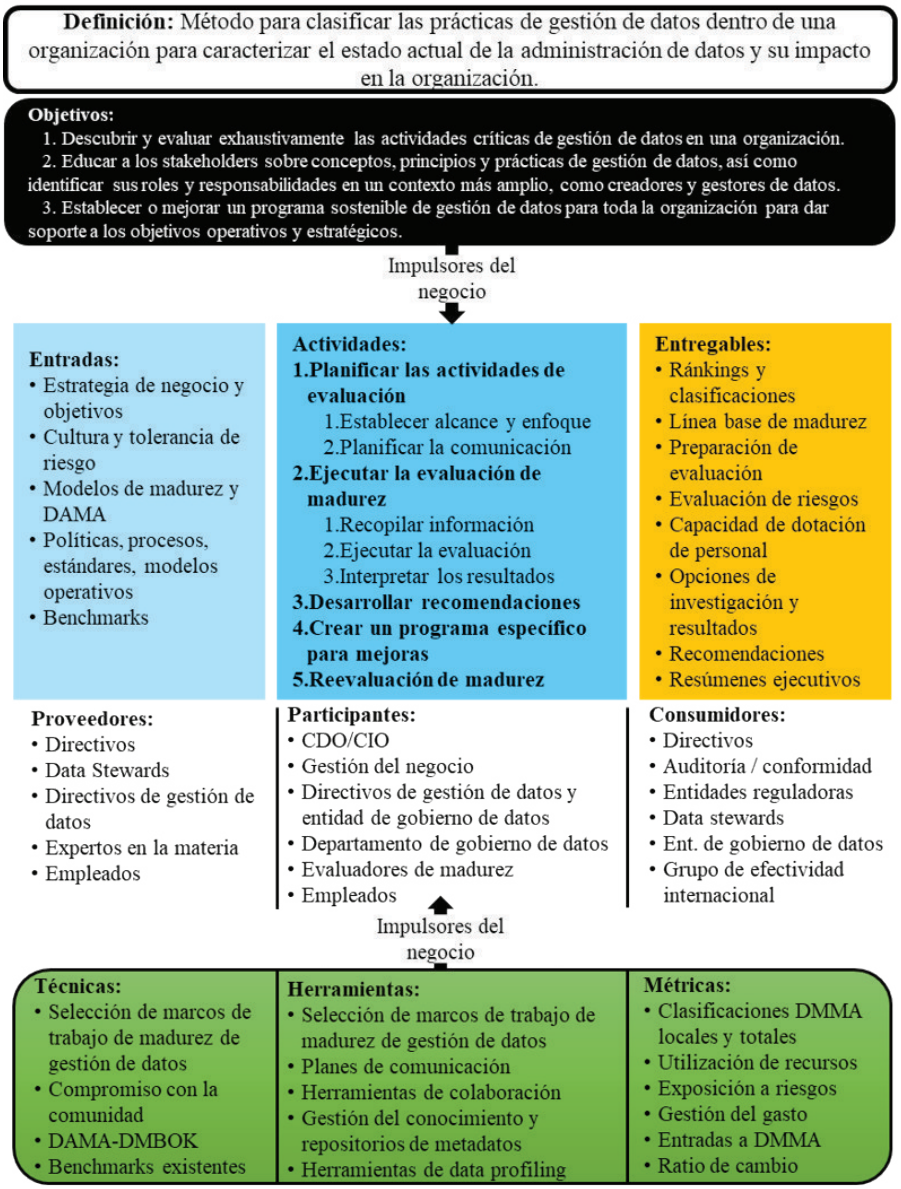

Figura 3.2. Esquema de proceso de gestión de datos del modelo (DAMA, 2017)

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

Y plantea como ejemplo un modelo de seis niveles:

- **Nivel 0 – Sin capacidades**. No existen prácticas de gestión de datos organizadas o procesos organizacionales formales para gestionar los datos. Muy pocas organizaciones suelen encontrarse en este nivel 0.

- **Nivel 1 – Inicial**. Se gestionan datos de propósito general usando un conjunto limitado de herramientas, con poco o ningún gobierno. El manejo de los datos depende en gran medida de unos pocos expertos. Los roles y responsabilidades se definen en "silos". Cada propietario de datos recibe, genera y envía datos de forma autónoma. Los controles, si existen, se aplican de forma inconsciente. Las soluciones para la gestión de datos son limitadas. Los problemas de calidad de datos son generalizados, pero no se abordan. El soporte a la infraestructura se realiza a nivel de unidad de negocio. Los criterios de evaluación pueden incluir la presencia de controles de proceso, como el registro de problemas de calidad de datos.

- **Nivel 2 – Repetible**. En este nivel, surge la implementación de herramientas consistentes y de definición de roles para el respaldo de la ejecución de los procesos. La organización comienza a utilizar herramientas centralizadas y a proporcionar más supervisión para la gestión de datos. Los roles se definen y los procesos no dependen únicamente de expertos específicos. Existe una conciencia organizacional de los problemas y conceptos de calidad de datos. También se reconocen los conceptos de datos maestros y de referencia. Los criterios de evaluación pueden incluir la definición formal de roles en artefactos como descripciones de trabajos, la existencia de documentación de procesos y la capacidad de aprovechar herramientas.

- **Nivel 3 – Definido.** Este nivel considera la introducción e institucionalización de procesos escalables de gestión de datos, que actúa como habilitador organizacional. Las características incluyen la replicación de datos en una organización con algunos controles implementados y un aumento general en la calidad general de los datos, junto con la definición y gestión coordinada de políticas. Una definición de proceso más formal conduce a una reducción significativa de la intervención manual. Esto, junto con un proceso de diseño centralizado, significa que los resultados del proceso son más predecibles. Los criterios de evaluación pueden incluir la existencia

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

de políticas de gestión de datos, el uso de procesos escalables y la coherencia de los modelos de datos y controles de sistemas.

- **Nivel 4 – Gestionado**. El conocimiento institucional obtenido del crecimiento en los niveles 1 a 3 permite a la organización predecir los resultados al abordar nuevos proyectos y tareas y comenzar a gestionar los riegos relacionados con los datos. La gestión de datos incluye métricas de rendimiento. Las características del nivel 4 incluyen herramientas estandarizadas para la gestión de datos junto con una función de gobierno y planificación centralizada. Las mejoras más notables de este nivel son un aumento medible en la calidad de los datos y en las capacidades de toda la organización. Los criterios de evaluación pueden incluir métricas relacionadas con el éxito del proyecto, las métricas operacionales para los sistemas y las métricas de calidad de los datos.

- **Nivel 5 – Optimizado.** Cuando las prácticas de gestión de datos están optimizadas, son altamente predecibles debido a la automatización de procesos y la gestión del cambio de tecnología. Las organizaciones en este nivel de madurez se enfocan en la mejora continua. En este nivel, las herramientas permiten ver datos en todos los procesos. La proliferación de datos se controla para evitar la duplicación innecesaria. Se usan métricas para gestionar y medir la calidad de los datos y los procesos. Los criterios de evaluación pueden incluir artefactos de gestión de cambios y métricas sobre la mejora de los procesos.

3.2 MODELO DE AIKEN

Aiken et al. (2007) plantean un modelo cuyo principal objetivo es incrementar los niveles de madurez de gestión de datos para conseguir un impacto positivo en la coordinación del flujo de datos entre organizaciones, recursos humanos y sistemas. Para mejorar las prácticas de gestión de datos de la organización, este modelo propone comenzar con una autoevaluación con respecto al nivel de madurez y elaborar una hoja de ruta para conseguir la mejora. El modelo establece que la gestión de datos consiste en seis procesos interrelacionados y coordinados (Véase Tabla 3.1).

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

Proceso	Descripción	Foco	Tipo de Datos
Programa de Coordinación de Datos	Proveer un proceso apropiado de gestión de datos y de infraestructura tecnológica	Dirección	Programa de datos: Proposiciones descriptivas u observaciones necesarias para establecer, documentar, sostener, controlar y mejorar las actividades organizacionales orientadas a datos (como visión, metas, políticas y métricas)
Integración Organizacional de Datos	Lograr el intercambio organizacional de datos apropiados	Dirección	Datos de desarrollo: Hechos descriptivos, proposiciones u observaciones usadas para desarrollar y documentar las estructuras e interrelaciones de los datos (por ejemplo, modelos de datos, diseños de bases de datos, y especificaciones)
Administración de Datos	Lograr la integración de datos del área temática del negocio	Dirección e Implementación	Administración de datos: Hechos descriptivos sobre documentación de la semántica y sintaxis de los datos (nombre, definición, formato…)
Desarrollo de los Datos	Lograr intercambio de datos dentro de un área de negocio	Implementación	Datos de negocio: Hechos y construcciones usados para lograr las actividades de negocio de la empresa (elemento de los datos, registros y archivos)
Datos de Soporte a Operaciones	Proveer un acceso confiable a los datos	Implementación	-
Uso Activo de los Datos	Aprovechar los datos en las actividades de negocio	Implementación	-

Tabla 3.1. Procesos de gestión de datos propuestos en Aiken et al. (2007).

Todas las organizaciones implementan sus prácticas de administración de datos de manera que pueden clasificarse en uno de los cinco niveles de modelo de madurez, detallados en la Tabla 3.2.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

Nivel	Nombre	Práctica	Calidad y resultados predecibles
1	Inicial	La organización carece de los procesos necesarios para sostener prácticas de gestión de datos. La gestión de datos está caracterizada por ser *ad hoc* o caótica.	La organización depende totalmente de los individuos, con más o menos visibilidad corporativa respecto al coste o rendimiento, o incluso tienen conciencia sobre prácticas de gestión de datos. Hay calidad variable, bajos resultados predecibles y poca o nula repetibilidad.
2	Repetible	La organización tiene cierto conocimiento sobre gestión de datos y puede replicar algunas buenas prácticas y casos de éxito.	La organización expone resultados con cierta calidad. El personal más cualificado es asignado a los proyectos críticos para reducir el riesgo y mejorar los resultados.
3	Definido	La organización utiliza un conjunto definido de procesos, que se publican para su uso.	Buenos resultados la mayor parte del tiempo.
4	Gestionado	La organización estadísticamente prevé y dirige la gestión de datos, basándose en procesos definidos, selección de coste, planificación, y niveles de satisfacción del cliente. Se requiere y monitoriza el uso de procesos de gestión de los datos dentro de la organización.	Resultados confiables y predecibles, así como la habilidad de determinar el progreso conseguido.
5	Optimizado	La organización analiza los procesos de gestión de datos existentes para determinar cuáles pueden ser mejorados, realizar cambios de una forma controlada y reduciendo costes operacionales mediante la mejora del rendimiento o introduciendo servicios innovadores para mantener su competitividad.	La organización alcanza altos niveles de resultados certeros.

Tabla 3.2. Niveles prácticos de medición de Gestión de datos propuestos en (Aiken et al., 2007).

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

3.3 DATA MANAGEMENT MATURITY MODEL (DMM)

El SEI (Software Engineering Institute) ha publicado recientemente el DMM (Data Management Maturity Model) (SEI, 2014), que es análogo al modelo de madurez para procesos software, CMMI (*Capability Maturity Model Integration*) pero centrado en los procesos de gobierno, gestión y calidad de datos, como se muestra en la Figura 3.3.

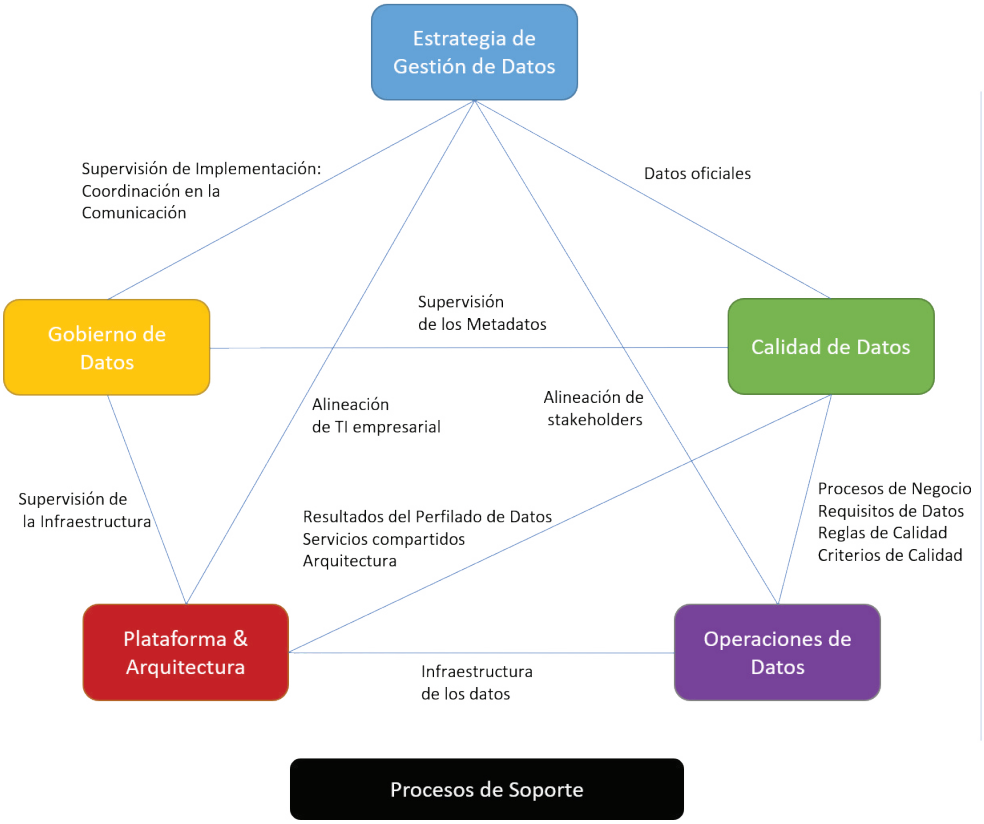

Figura 3.3. Modelo DMM (SEI, 2014).

El modelo DMM se compone de 6 categorías de procesos que agrupan 25 procesos (véase Tabla 3.3).

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

DMM	
Estrategia de Gestión de Datos	DMS 1.1 Estrategia de gestión de datos
	DMS 1.2 Comunicaciones
	DMS 1.3 Función de gestión de datos
	DMS 1.4 Caso de negocio
	DMS 1.5 Programa de financiación
Gobierno de Datos	DG 2.1 Gestión del gobierno
	DG 2.2 Glosario de negocio
	DG 2.3 Gestión de metadatos
Calidad de Datos	DQ 3.1 Estrategia de calidad de datos
	DQ 3.2 Perfilado de datos
	DQ 3.3 Evaluación de calidad de datos
	DQ 3.4 Limpieza de datos
Operaciones de Datos	DO 4.1 Definición de requisitos de datos
	DO 4.2 Gestión del ciclo de vida de los datos
	DO 4.3 Gestión de proveedores
Plataforma y Arquitectura	PA 5.1 Enfoque de arquitectura
	PA 5.2 Estándares de arquitectura
	PA 5.3 Plataforma de gestión de datos
	PA 5.4 Integración de datos
	PA 5.5 Datos históricos, archivamiento y retención
Procesos de Soporte	SP 6.1 Medición y análisis
	SP 6.2 Gestión de procesos
	SP 6.3 Aseguramiento de calidad de procesos
	SP 6.4 Gestión del riesgo
	SP 6.5 Gestión de la configuración

Tabla 3.3. Modelo de referencia de procesos de DMM.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

3.3.1 Estrategia de Gestión de Datos

Los procesos correspondientes a la estrategia de Gestión de datos son los siguientes:

DMS 1.1 Estrategia de gestión de datos

Define la visión, metas y objetivos para el programa de gestión de datos y asegura que todos los *stakeholder*s relevantes estén alineados con las prioridades y la implementación y gestión de los programas.

DMS 1.2 Comunicaciones

Asegura que se publican, documentan, entienden y ajustan -basándose en los comentarios de los *stakeholders*-, las políticas, estándares, procesos, anuncios de progreso y otras comunicaciones relacionadas con la gestión de datos.

DMS 1.3 Función de gestión de datos

Ofrece una guía para el líder y el personal de gestión de datos para asegurar que los datos se gestionan como un activo corporativo. La supervisión de la dirección es crítica para establecer y mantener los principios de gestión de datos, facilitar la adopción y asegurar el alineamiento de toda la organización.

DMS 1.4 Caso de negocio

Da la razón fundamental para determinar qué iniciativas de gestión de datos deben ser financiadas y asegura la sostenibilidad de la gestión de datos tomando decisiones basadas en consideraciones financieras y en beneficios para la organización.

DMS 1.5 Programa de financiación

Asegura la disponibilidad de financiación adecuada y sostenible para apoyar el programa de gestión de datos.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

3.3.2 Gobierno de Datos

Los procesos que DMM propone para el Gobierno de Datos son los siguientes:

DG 2.1 Gestión del gobierno

Desarrolla la estructura necesaria para asegurar que los datos corporativos se gestionan como un activo crítico y se implementan de forma efectiva y sostenible.

DG 2.2 Glosario de negocio

Contribuye a un entendimiento común de términos y definiciones sobre los datos estructurados y no estructurados, que soportan los procesos de negocio para todos los *stakeholders*.

DG 2.3 Gestión de metadatos

Establece los procesos y la infraestructura para especificar y extender información clara y organizada sobre los activos de datos estructurados y no estructurados bajo gestión, fomentando y soportando la compartición de datos, asegurando el uso adecuado de los datos, mejorando la capacidad de respuesta a los cambios en el negocio y reduciendo los riesgos relacionados con los datos.

3.3.3 Calidad de Datos

Con respecto a la calidad de los datos, los procesos que se proponen en DMM, son:

DQ 3.1 Estrategia de calidad de datos

Define una estrategia integrada para toda la organización para conseguir y mantener el nivel de calidad de datos requerido para soportar las metas y objetivos del negocio.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

DQ 3.2 Perfilado de datos

Desarrolla una comprensión del contenido, calidad y reglas de un específico conjunto de datos bajo gestión.

DQ 3.3 Evaluación de calidad de datos

Ofrece una propuesta para medir y evaluar de forma sistemática la calidad de los datos de acuerdo con procesos, técnicas y reglas de calidad de datos.

DQ 3.4 Limpieza de datos

Define los mecanismos, reglas, procesos y métodos usados para validar y corregir los datos de acuerdo con reglas de negocio predefinidas.

3.3.4 Operaciones de Datos

Los procesos descritos por DMM correspondientes a las operaciones de datos son los siguientes:

DO 4.1 Definición de requisitos de datos

Asegura que los datos producidos y consumidos satisfacen los objetivos de negocio, que se comprenden por todos los *stakeholders* relevantes y que son consistentes con los procesos que crean y consumen los datos.

DO 4.2 Gestión del ciclo de vida de los datos

Asegura que la organización comprende, relaciona, inventaría y controla sus flujos de datos a través de los procesos de negocio a lo largo de todo el ciclo de vida de los datos desde su creación o adquisición hasta su retirada. La gestión del ciclo de vida de los datos habilita una mejor gestión de riesgos y soporta mejoras a la calidad de datos, particularmente en situaciones que involucran a grandes volúmenes de datos o con una alta velocidad de transferencia, y en procesos complejos e interdependientes que comparten datos.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

DO 4.3 Gestión de proveedores

Optimiza las fuentes de datos internas y externas para satisfacer los requisitos de negocio y gestionar los acuerdos de aprovisionamiento de datos de forma consistente.

3.3.5 Plataforma y Arquitectura

En lo que respecta a la Plataforma y Arquitectura, los procesos que incorporar DMM son los siguientes:

PA 5.1 Enfoque de arquitectura

Diseña e implementa una capa óptima de datos que habilita la adquisición, producción, almacenamiento y entrega de datos para cumplir los objetivos técnicos y de negocio.

PA 5.2 Estándares de arquitectura

Ofrece un conjunto aprobado de expectativas para gobernar elementos arquitecturales que soportan representaciones de datos aprobadas, acceso a los datos y distribución de datos, fundamentales para el control de los activos de datos y el uso e intercambio eficientes de información.

PA 5.3 Plataforma de gestión de datos

Asegura que se implementa y se gestiona una plataforma efectiva para satisfacer las necesidades de negocio.

PA 5.4 Integración de datos

Reduce la necesidad del negocio de obtener datos de múltiples fuentes y de mejorar la disponibilidad de los datos para los procesos de negocio que requieren consolidación y agregación de datos. La integración de datos habilita la optimización de fuentes de datos, la reducción de costes a través de la centralización, y la mejora de la calidad de los datos.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

PA 5.5 Datos históricos, archivado y retención

Asegura que el mantenimiento de los datos satisface los requisitos organizacionales y normativos para la disponibilidad de datos históricos, y los requisitos legales y normativos para el archivado y la conservación de datos.

3.3.6 Procesos de Soporte

Finalmente, los procesos de soporte que incorpora DMM son:

SP 6.1 Medición y análisis

Desarrolla y sostiene la capacidad de medición y las técnicas analíticas para apoyar la gestión y mejora de las actividades de gestión de datos.

SP 6.2 Gestión de procesos

Establece y mantiene un conjunto usable de activos organizacionales de proceso y planifica, implementa y despliega mejoras de los procesos organizacionales dirigidos por los objetivos de negocio y las brechas actuales en los procesos de la organización.

SP 6.3 Aseguramiento de calidad de procesos

Proporciona personal y gestión con una visión objetiva de la ejecución del proceso y los productos de trabajo asociados.

SP 6.4 Gestión del riesgo

Identifica y analiza problemas potenciales para tomar las acciones apropiadas con el fin de asegurar que se puedan conseguir los objetivos.

SP 6.5 Gestión de la configuración

Establece y mantiene la integridad del entorno operacional usando identificación, control, monitorización del estado y auditoría de la configuración.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

3.3.7 Niveles de Madurez en DMM

DMM presenta cinco niveles de madurez, que se resumen en la Tabla 3.4.

Nivel	Descripción	Perspectiva
1: Realizado	Los procesos se realizan *ad hoc*, principalmente a nivel de proyecto. La disciplina del proceso es reactiva; por ejemplo, los procesos de calidad de datos enfatizan la reparación sobre la prevención. Deben existir mejoras fundamentales, pero las mejoras aún no se extienden a toda la organización ni se mantienen.	Los datos se gestionan como un requisito para la implementación de proyectos.
2: Gestionado	Los procesos se planifican y se ejecutan de acuerdo con las políticas; se emplean personas cualificadas con recursos adecuados para producir resultados controlados; se involucra a los *stakeholders* más importantes; se monitorizan, controlan y evalúan para determinar el nivel de adherencia al proceso definido.	Existe preocupación por la importancia de la gestión de los datos como un activo crítico infraestructural.
3: Definido	Se emplea un conjunto de procesos estándar y se siguen consistentemente. Los procesos para satisfacer necesidades específicas se adaptan a partir del conjunto de procesos estándar de acuerdo con las directrices de la organización.	Los datos se tratan a nivel organizacional como críticos para el desempeño exitoso de la misión de la organización.
4: Medido	Se han definido métricas de proceso y se usan para la gestión de datos, incluyendo la gestión de la varianza, la predicción y el análisis utilizando técnicas estadísticas y técnicas cuantitativas. El rendimiento del proceso se gestiona a lo largo de la vida del proceso.	Los datos se tratan como una fuente de ventaja competitiva.
5: Optimizado	El rendimiento del proceso se optimiza mediante la aplicación de los análisis de nivel 4 para la identificación de oportunidades de mejora. Las mejores prácticas se comparten con la industria.	Se perciben los datos como algo crítico para la supervivencia en un mercado dinámico y competitivo.

Tabla 3.4. Modelo de madurez de DMM.

3.4 MODELO DE IBM

IBM (Soares, 2010) propone una guía para implantar un programa de gobierno de datos, que se evalúa con respecto a un modelo de madurez de gobierno de datos. A continuación, se detallan ambos.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

3.4.1 Proceso unificado de gobierno de datos de IBM

El proceso unificado de gobierno de datos de IBM propone catorce pasos que se muestran en la Figura 3.4 (diez pasos obligatorios y cuatro caminos opcionales) para realizar un programa de gobierno de datos efectivo.

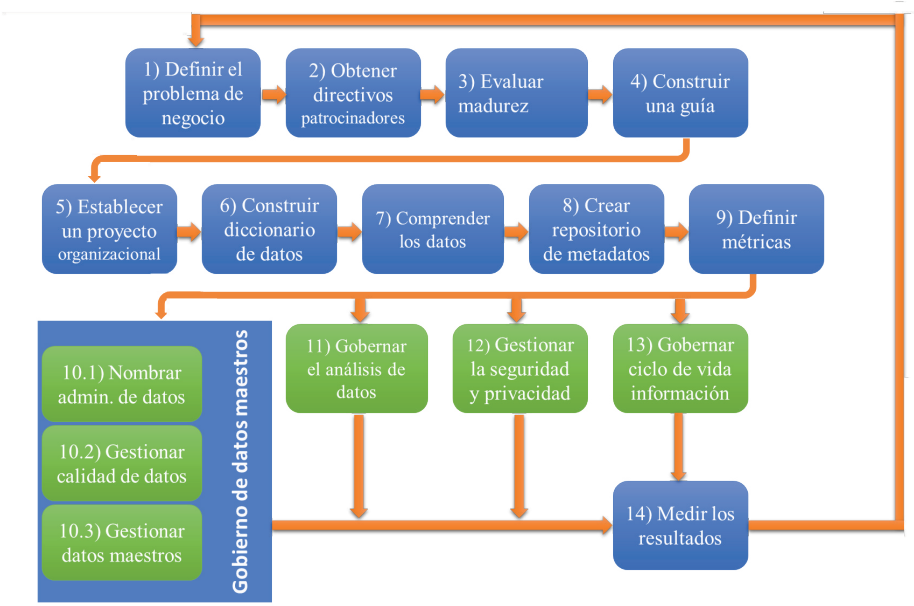

Figura 3.4. Proceso unificado de gobierno de datos IBM (Adaptado de Soares, 2011)

Los diez pasos obligatorios (del 1 al 9 y el 14) son necesarios para sentar las bases de un programa de gobierno de datos eficaz. En cuanto a los caminos opcionales, una organización seleccionará uno o más: gobierno de datos maestros, gobierno de analíticas, gestionar la seguridad y privacidad y gobernar el ciclo de vida de la información. Finalmente, se necesita medir el proceso y transmitir los resultados a los directivos patrocinadores de forma regular. A continuación, se describen los pasos del modelo en mayor detalle (Soares, 2010):

1. **Definir el problema de negocio.**

   Es imprescindible que la organización defina el alcance inicial del programa de gobierno de datos sobre un problema de negocio específico, como una auditoría fallida, una violación en los datos o la necesidad de mejorar la calidad de datos para propósitos de gestión de riesgos.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

Una vez el programa empieza a abordar los problemas identificados, recibirá apoyo para extender su alcance a áreas adicionales.

2. **Obtener directivos patrocinadores.**

   Es importante establecer directivos patrocinadores de TI y de negocio para el programa de gobierno de datos. La mejor forma de obtener el patrocinio es establecer valor en términos de un caso de negocio y plantear objetivos rápidos.

3. **Llevar a cabo una evaluación de madurez.**

   Cada organización necesita llevar a cabo una evaluación de su madurez en gobierno de datos, preferiblemente de forma anual. Para ello, IBM propone un modelo de madurez organizacional (véase apartado 3.4.2) de gobierno de datos basado en once categorías que se detalla en el siguiente apartado. La organización necesita evaluar el nivel de madurez y fijar un nivel de madurez futuro para conseguir alcanzarlo en doce a dieciocho meses.

4. **Construir una guía.**

   La organización necesita desarrollar una guía para implantar las necesidades existentes entre el estado actual y el estado futuro deseado, para las once categorías de madurez de gobierno de datos.

5. **Establecer un diseño arquitectónico organizacional.**

   La organización de gobierno de datos necesita un estatuto para gobernar sus operaciones y asegurar que tiene suficiente autoridad para decidir en situaciones críticas.

6. **Construir el diccionario de datos.**

   Se debe construir un diccionario de datos común, que podrá abarcar la organización para asegurar que los términos de negocio están enlazados a través de los metadatos a los términos técnicos y que la organización tiene una única comprensión común de los datos.

7. **Comprender los datos.**

   El equipo de gobierno de datos necesita descubrir las relaciones de datos críticas en la organización. El descubrimiento de datos puede incluir relaciones simples y otras más complejas de encontrar, así como

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

las ubicaciones de datos sensibles en los sistemas de información de la organización.

8. **Crear un repositorio de metadatos.**

   Los metadatos se deben almacenar en un repositorio para que puedan ser compartidos en múltiples proyectos.

9. **Definir métricas.**

   El gobierno de datos debe disponer de métricas robustas para medir y controlar el progreso. El equipo de gobierno de datos debe elegir indicadores (KPIs – *Key Performance Indicators*) para medir el rendimiento del programa.

Estos son los primeros nueve pasos obligatorios, en este punto la organización puede elegir uno o más de los siguientes pasos opcionales

10. **Gobernar los datos maestros.**

    La información más valiosa para una organización suele ser conocida como datos maestros. Gobernar los datos maestros es una práctica continua, donde los líderes de negocio definen principios, políticas, procesos, reglas de negocio y métricas para conseguir los objetivos de negocio gestionando la calidad de sus datos maestros.

11. **Gobernar el análisis (*analytics*) de datos.**

    Se considera el gobierno del análisis de los datos como el conjunto de políticas y procedimientos para alinear mejor a los usuarios del negocio con las inversiones en la infraestructura analítica.

12. **Gestionar la seguridad y la privacidad.**

    Los responsables de gobierno de datos, especialmente los que están relacionados directamente con el responsable de seguridad de la información, tienen que tratar los problemas de seguridad y privacidad de datos.

13. **Gobernar el ciclo de vida de la información.**

    Las organizaciones deben tratar todos los problemas que se puedan generar a lo largo del ciclo de vida de la información.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

Y, por último, el último paso obligatorio:

14. **Medir los resultados.**

    El gobierno de datos debe asegurar la mejora continua, monitorizando constantemente las métricas En el paso nueve (definir métricas), el equipo de gobierno de datos establece las métricas, mientras que, en este paso, el equipo informa sobre el progreso en relación con esas métricas a los principales stakeholders de TI y de negocio.

3.4.2 Modelo de madurez de gobierno de datos de IBM

El modelo de madurez de gobierno de datos de IBM ha sido desarrollado por el IBM *Data Governance Council* y está enfocado a ayudar a que la estrategia sea más efectiva. El modelo de madurez define el alcance y quién debe participar en el gobierno y la medición de la forma en que las organizaciones gobiernan sus datos. Este modelo mide las competencias de gobierno de datos basado en once categorías de madurez que se muestran en la Figura 3.5 (Soares, 2010).

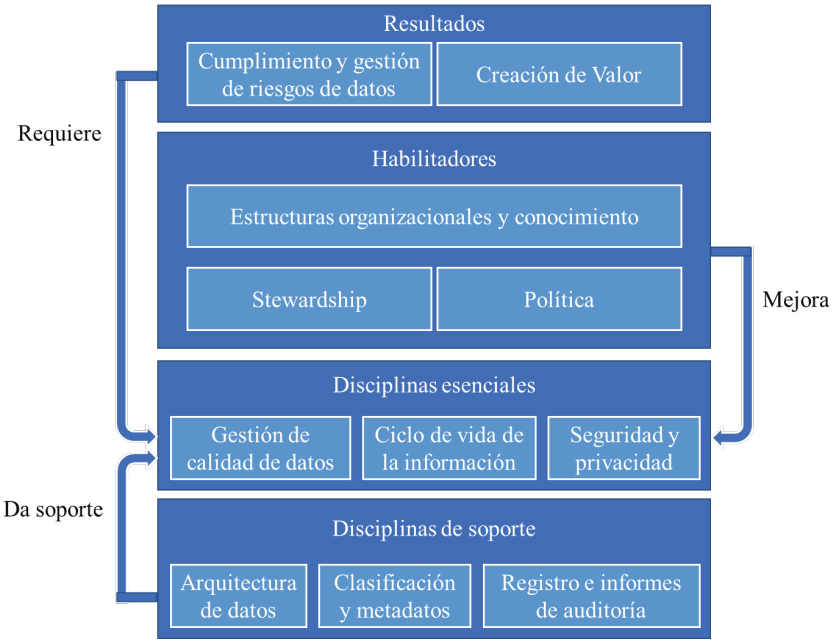

Figura 3.5. Modelo de madurez de gobierno de datos de IBM.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

El modelo de madurez consta de cuatro grupos interrelacionados:

- **Resultados.**

  Son los resultados previstos del programa de gobierno de datos, que tienden a centrarse en reducir riesgos e incrementar el valor, y que a su vez son impulsados por la reducción de costes y el aumento de los ingresos.

- **Habilitadores.**

  Incluye áreas de las estructuras organizacionales y conocimiento, políticas y custodia de datos (*stewardship*).

- **Disciplinas esenciales.**

  Incluye la gestión de calidad de datos, gestión del ciclo de vida de datos, y seguridad y privacidad de datos.

- **Disciplinas de soporte.**

  Incluye la arquitectura de datos, clasificación y metadatos, y registro e informes de auditoría.

En cada uno de estos grupos se encuentran las siguiente once categorías:

1. **Cumplimiento y gestión de riesgo de datos.** Es una metodología en la que se identifican, se califican, se cuantifican, se aceptan, se evitan, se mitigan o se transfieren los riesgos.

2. **Creación de valor.** Es un proceso por el cual se califican y se cuantifican los activos de datos para maximizar el valor creado por los activos de datos.

3. **Estructuras organizacionales y conocimiento.** Se refiere al nivel de responsabilidad mutua entre el negocio y las TI y el reconocimiento de la responsabilidad fiduciaria para gobernar los datos en los diferentes niveles de gestión.

4. ***Stewardship*.** Es una disciplina de control de calidad diseñada para asegurar la custodia de los datos para la mejora de los activos, mitigación de riesgos y el control organizacional.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

5. **Política.** Es la articulación escrita del funcionamiento organizacional.
6. **Gestión de calidad de datos.** Se refiere a métodos para medir, mejorar y certificar la calidad e integridad de los datos de producción, de prueba y de archivo.
7. **Gestión del ciclo de vida de la información.**

como esencial para la madurez de la gestión de la información: visión, estrategia, métricas, gobierno, personas, procesos e infraestructura (Laney, 2017).

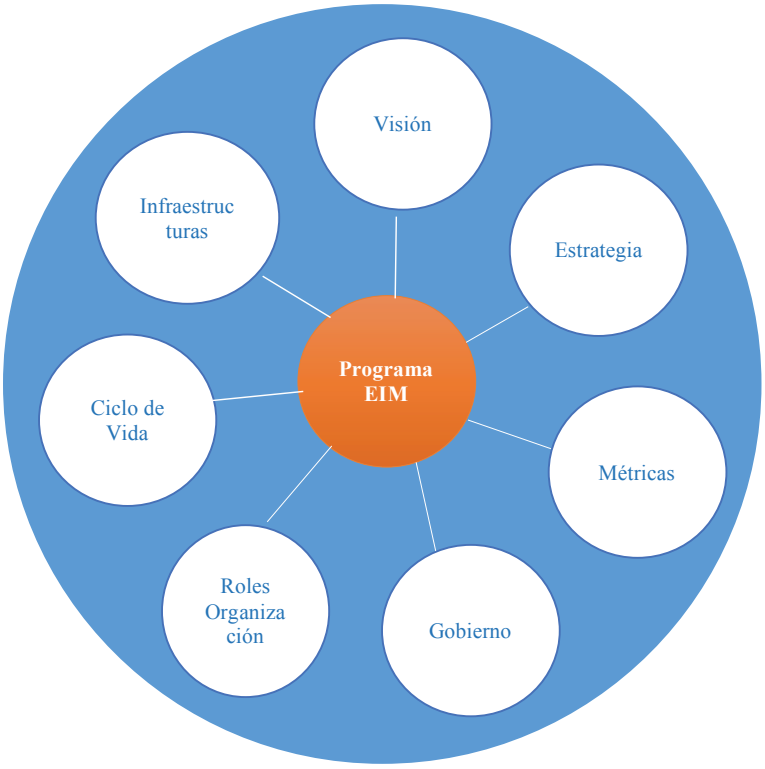

Figura 3.6. Modelo de madurez de gestión de la información de Gartner (Laney, 2017).

Los niveles de madurez e indicadores en sí mismos están alineados con las capacidades actuales y a corto plazo de las organizaciones:

- **Nivel 1**: Las organizaciones son conscientes de los problemas clave y de los cambios, pero carecen de recursos, presupuestos, y/o liderazgo para afrontar o realizar cambios significantes en el EIM.
- **Nivel 2**: Las organizaciones trabajan reactivamente centrados en las aplicaciones (*application-centric*), hasta que los problemas relacionados con la información se manifiestan significativamente en

- **Nivel 3**: Las organizaciones se han convertido en organizaciones más proactivas en identificar ciertas áreas de gestión de información, y han empezado a identificar la "organización" en los sistemas de información. Algunos programas son operativos y efectivos, pero hay poca influencia o alineación entre los programas y las inversiones.

- **Nivel 4:** Adoptan un enfoque administrado para la gestión de la información, comprometiendo la coordinación en toda la organización, con personas, procesos y tecnologías eficaces.

- **Nivel 5**: Normalmente son organizaciones modelo en las que se han optimizado muchos (si no la mayoría) de aspectos de la adquisición, administración y aplicación de información como un activo organizacional real con estructuras de organización de alto rendimiento y uso de tecnologías y arquitecturas avanzadas.

3.6 TQDM

Larry English (English, 1999) propuso la metodología TQdM (*Total Quality data Management*) que se traduce en la mejora continua de dos categorías de procesos:

- Procesos de desarrollo de sistemas de información para definir información, desarrollando e implementando procesos de negocio, sistema de información, y arquitectura de información y bases de datos.

- Procesos de negocios que crean, actualizan y borran datos, distribuyen o reparten información y recuperan o presentan información a los productores de información o a los trabajadores del conocimiento. TQdM busca conseguir estos objetivos integrando los principios y

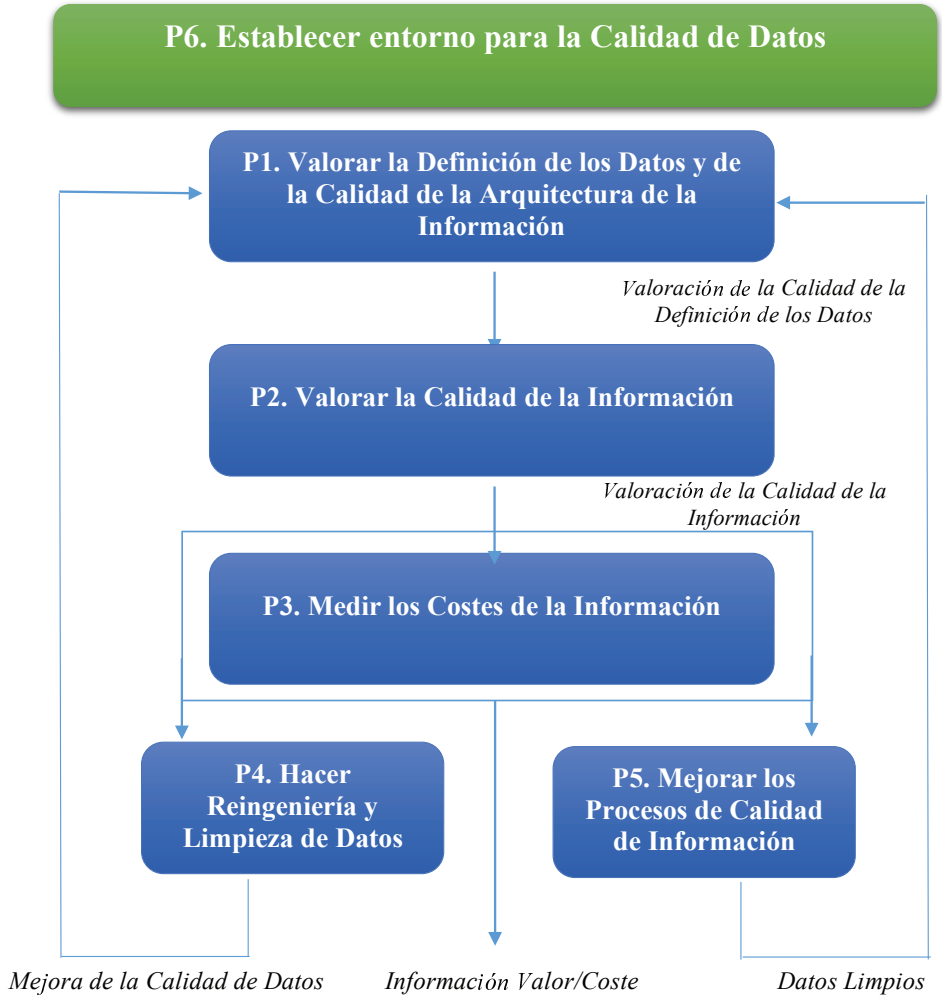

Figura 3.7. Metodología TQdM (English, 1999).

3.7 DCAM

El DCAM (*Data Management Capability Assessment Model*)[4] fue creado por los miembros del Consejo EDM (Enterprise Data Management) como un conjunto

de los datos. DCAM documenta 36 capacidades y 115 sub-capacidades asociadas con el desarrollo de un programa sostenible de gestión de datos.

En la Figura 3.8 se puede observar la división de las capacidades en sub-capacidades, objetivos y artefactos. T

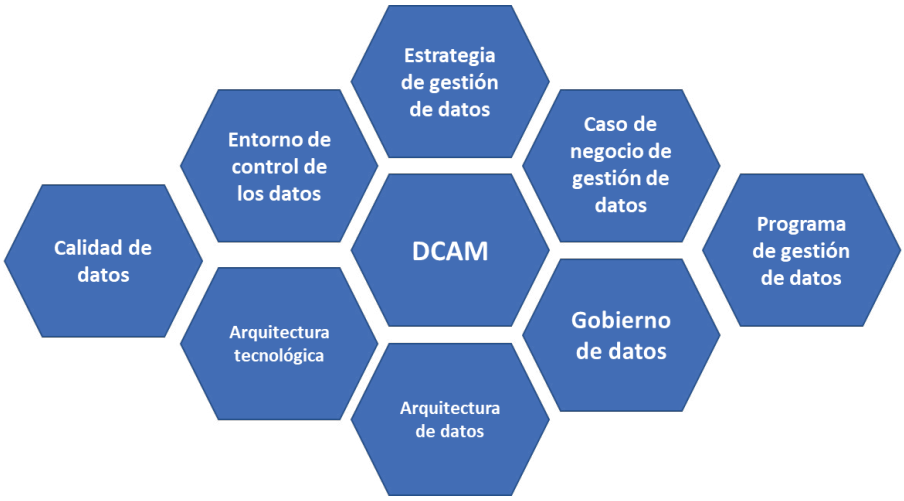

Figura 3.9. Componentes del modelo DCAM.

Como se puede ver en la Figura 3.9, hay 8 componentes y cada componente tiene distintas capacidades. En total hay 36 capacidades, 112 sub-capacidades y

3.8.1 Visión general del modelo

El modelo MAMD cubre las tres disciplinas relacionadas con los datos: gestión de datos, gestión de calidad de datos y gobierno de datos, con el objetivo de que trabajen de forma integrada.

Ventajas a nivel operativo

- Tener identificadas formalmente responsabilidades respecto a los datos que gestiona y utiliza.
- Disponer de directivas para manejar los datos asegurando los niveles de calidad de datos.
- Recibir formación

3.8.2.1 PROCESOS DE GESTIÓN DE DATOS

DM.1. Gestión de requisitos de datos

Recopilar y validar los requisitos referentes a los datos necesarios para cumplir los objetivos organizacionales.

DM.2. Gestión de la infraestructura tecnológica

Proporcionar infraestructura y servicios de datos para satisfacer los objetivos de negocio.

DM.3. Gestión de datos históricos

Archivar y almacenar los datos satisfaciendo los requisitos legislativos, normativos y organizacionales.

DM.4. Gestión de seguridad de datos

Definir y habilitar mecanismos para hacer posible la confidencialidad, integridad y disponibilidad de los datos.

DM.5. Gestión de la configuración de datos

Establecer y mantener la integridad de todos los activos relacionados con

DM.8. Establecimiento de fuentes y destinos de datos

Identificar y caracterizar cada una de las fuentes y de los destinos de los datos usados en los procesos de negocio organizacionales, así como los acuerdos y las interacciones con los proveedores y clientes.

DM.9. Integración de datos

Asegurar la consolidación y agregación de los datos mediante el control de flujo y las relaciones con los datos transferidos a los sistemas de información.

3.8.2.2 PROCESOS DE GESTIÓN DE LA CALIDAD DE LOS DATOS

DQM.1. Planificación de calidad de datos

Establecer los recursos necesarios para satisfacer los requisitos, y medir los niveles de calidad de datos de acuerdo con los criterios de medición.

DQM.2. Monitorización y control de calidad de datos

Realizar una medición continua de la calidad de datos en diferentes puntos de los procesos de negocio para determinar si los datos satisfacen los niveles de calidad adecuados.

DQM.3. Aseguramiento de calidad de datos

Realizar una medición de la ejecución de los procesos relacionados con las no conformidades de los datos encontradas en la monitorización y control y evaluar el impacto.

DQM.4. Mejora de calidad de datos

Analizar las causas de los problemas de calidad de datos para prevenir futuras no conformidades y corregirlos.

3.8.2.3 PROCESOS DE GOBIERNO DE DATOS

DG.1. Establecimiento de estrategias de datos

Definir objetivos para mantener un nivel adecuado de calidad de datos

DG.2. Gestión del ciclo de vida de los datos

Controlar el uso de los datos en los procesos de negocio desde su creación o adquisición hasta la eliminación.

DG.3. Gestión del valor de los datos

Identificar el grado de importancia que tienen los datos para los diferentes procesos de negocio.

DG.4. Establecimiento de estándares, políticas, buenas prácticas y procedimientos

Establecer aquellos

3.8.2.4 RESUMEN DEL MODELO DE REFERENCIA DE PROCESOS

En la Tabla 3.6 se puede ver a modo de resumen el modelo de referencia de procesos con la identificación de los procesos que pertenecen a cada disciplina: gestión de datos (DM), gestión de la calidad de los datos (DQM) y gobierno de datos (DG).

Área	Proceso
DM	DM.1. Gestión de Requisitos de Datos
	DM.2. Gestión de la Infraestructura Tecnológica
	DM.3. Gestión de Datos Históricos
	DM.4. Gestión de la Seguridad de Datos
	DM.5. Gestión de la Configuración de Datos
	DM.6. Gestión de Datos Maestros y de Referencia
	DM.7. Arquitectura y Diseño de Datos
	DM.8. Establecimiento de Fuentes y Destinos de Datos
	DM.9. Integración de Datos
DQM	DQM.1. Planificación de Calidad de Datos
	DQM.2. Monitorización y Control de Calidad de Datos
	DQM.3. Aseguramiento de Calidad de datos
	DQM.4. Mejora de Calidad de Datos
DG	DG.1. Establecimiento de Estrategias de Datos
	DG.2. Gestión del Ciclo de Vida de los Datos
	DG.3. Gestión del Valor de los Datos
	DG.4. Establecimiento de Estándares, Políticas y Buenas Prácticas
	DG.5. Gestión de Recursos Humanos
	DG.6. Gestión de Recursos Financieros
	DG.7. Monitorización de las Estrategias Organizacionales de Datos
	DG.8. Gestión de Cambios en las Estrategias de Datos

Tabla 3.6. Modelo de referencia de procesos de MAMD.

3.8.3 Modelo de evaluación

El modelo de evaluación de MAMD incorporar un modelo de madurez basado en la capacidad de los procesos incluidos en el modelo de referencias de procesos.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

Como se describe en ISO/IEC 33020 (ISO, 2015d), la capacidad de un proceso es un valor que se escoge de una en una escala ordinal de seis puntos que habilitan la evaluación de la capacidad desde el nivel inferior, incompleto, hasta el m s alto, innovando. Para calcular el nivel de capacidad de los procesos es necesario observar las evidencias para cuantificar el nivel de implantación de los requisitos descritos para cada uno de los atributos de procesos. La Tabla 3.7 introduce los atributos de proceso y los niveles de capacidad que tienen que conseguirse.

Nivel de capacidad del proceso	Atributos de proceso
Incompleto	No Aplica
Realizado	AP.1.1 Proceso realizado
Gestionado	AP.2.1 Gestión AP.2.2 Gestión de productos de trabajo
Establecido	AP.3.1 Definición del proceso AP.3.2. Despliegue del proceso
Predecible	AP.4.1 An lisis cuantitativo AP.4.2 Control cuantitativo
Innovando	AP.5.1 Innovación del proceso AP.5.2 Implementación de la innovación del proceso

Tabla 3.7. Niveles de capacidad y atributos de proceso.

Como se decía anteriormente, para determinar el nivel de capacidad de los procesos del modelo de referencia de procesos de MAMD hay que realizar una inspección de los distintos tipos de evidencias (directas, indirectas, entrevistas y documentación) que se recolectarán para cada una de las instancias de los procesos de negocio que se deben escoger para realizar la evaluación. Han de escogerse un proceso principal y tres procesos auxiliares en los que buscar las evidencias de la evaluación de los procesos de datos que se

Ordinal	Significado
N – No satisfecho	Hay pocas o ninguna evidencia de que los requisitos de los atributos de proceso han sido implementado mínimamente para los procesos evaluados.
P – Parcialmente satisfecho	Hay evidencias d biles de que se han implementado m s o menos los requisitos de los atributos de proceso definidos para los procesos de MAMD. No obstante, algunos aspectos de la consecución de los atributos de procesos pueden ser impredecibles.
L – Ampliamente satisfecho	Hay evidencias fuertes y significativas de la implementación de los requisitos para los atributos de proceso definidos en los procesos evaluados. Se pueden encontrar algunas debilidades menores en el atributo de proceso del proceso evaluado.
F – Totalmente satisfecho	Las evidencias son fuertes y reflejan una completa y sistem tica realización total de los requisitos de los atributos de proceso definidos en el proceso de evaluación. No se encuentran debilidades relacionadas con este atributo de proceso.

Tabla 3.8. Escala ordinal para la calificación de los atributos de proceso (ver ISO/IEC 33020)

Para poder afirmar que un proceso del modelo de referencia de MAMD ha alcanzado un determinado nivel de capacidad todos los atributos de proceso de ese nivel de capacidad deben tener al menos una calificación de *Ampliamente satisfecho (L) o Totalmente satisfecho (F).* Adem s, es necesario que todos sus predecesores tengan una calificación de *Totalmente satisfecho* en cada atributo de proceso.

Para poder determinar si se ha alcanzado un nivel de madurez hay que evaluar el nivel de capacidad de todos los procesos que pertenecen a ese nivel de madurez según se registra en el modelo de madurez. La siguiente sección introduce el modelo de madurez.

3.8.4 Modelo de madurez

Es importante tener en cuenta que MAMD est alineado a ISO 8000-62. La clasificación de los procesos del modelo de referencia de procesos en cada uno de los niveles del modelo de madurez se ha realizado teniendo en cuenta distintos criterios: los beneficios que las buenas prácticas de los procesos identificados pueden suponer para maximizar el rendimiento de los procesos de negocio si se implantan de forma temprana,,cómo otros modelos de madurez existentes en el panorama han asignado estos procesos a distintos niveles, de los procesos en otros modelos de referencia, la complejidad del procesos, y los recursos necesarios que la organización debe invertir o tener disponibles para tener niveles adecuados de calidad de datos.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

Los niveles de madurez que se proponen en MAMD, junto con su significado y los procesos que se incluyen, se detallan a continuación.

Inmaduro

La organización no puede demostrar la implementación efectiva de las buenas pr cticas del modelo de referencia de procesos. Por lo tanto, no hay garantías de que estén gestionando y utilizando adecuadamente sus datos.

Básico

La organización puede demostrar que utiliza un conjunto de buenas pr cticas orientadas a proporcionar el soporte mínimo necesario para la gestión de los datos necesarios en sus procesos de negocio. Pero no se presta en absoluto ninguna atención ni al gobierno de los datos ni a la calidad de datos.

Gestionado

La organización puede demostrar que utiliza un conjunto de buenas pr cticas orientadas a garantizar que los datos usados en sus procesos de negocio est n alineados con la estrategia organizacional. Por lo tanto, hay garantías de que la organización tiene implementado los procesos mínimos necesarios de gobierno de datos como para asegurar el xito en sus procesos de negocio. Esto implica que los siguientes procesos han alcanzado el nivel de capacidad 2 con una valoración de al menos *ampliamente* para los atributos de procesos AP.2.1 y AP.2.2:

- DM.1. Gestión de requisitos de datos.
- DM.2. Gestión de infraestructura tecnológica.
- DM.3. Gestión de datos históricos.
- DM.4. Gestión de seguridad de datos.
- DM.5. Gestión de la configuración.
- DQM.2. Monitorización y control de calidad de datos.
- DG.4. Establecimiento de estándares, políticas, buenas pr cticas y procedimientos.

Establecido

La organización puede demostrar que utiliza un conjunto de buenas pr cticas orientadas a gestionar la calidad de los datos para garantizar que los

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

datos usados en sus procesos de negocio tienen los niveles adecuados de calidad. Esto implica que los procesos de negocio se benefician de la implantación de los requisitos establecidos en los atributos de procesos AP.3.1 y AP.3.2 no sólo de los procesos listado para el nivel 2 (que deben tener un nivel de capacidad con una calificación de *completamente satisfecho*) sino adem s de los siguientes (que pueden tener un nivel de capacidad con los atributos de procesos AP.3.1 y AP.3.2 con una valoración de al menos *ampliamente satisfecho*:

- DM.6. Gestión de datos maestros
- DM.7. Diseño de datos
- DM.8. Establecimiento de fuentes y destinos de datos.
- DM.9. Integración de datos
- DQM.1. Planificación de Calidad de Datos
- DG.1. Establecimiento de estrategias de datos
- DG.2. Gestión del ciclo de vida
- DG.5. Gestión de recursos humanos

Predecible

La organización puede demostrar que utiliza un conjunto de buenas pr cticas orientadas a monitorizar que las estrategias organizacionales de datos son realmente efectivas. Para alcanzar este nivel una organización debe tener todos los procesos del nivel de madurez 3 con una calificación de al menos "*completamente satisfecho",* y asegurarse de que los procesos del nivel 4 est n al menos con una calificación de *ampliamente satisfechos*:

- DQM.3. Aseguramiento de Calidad de Datos.
- DG.3. Gestión del valor de los datos.
- DG.6. Gestión de recursos financieros.
- DG.7. Monitorización de las estrategias organizaciones de datos.

Innovando

La organización puede demostrar que utiliza un conjunto de buenas pr cticas orientadas a garantizar que las estrategias de datos van evolucionando a medida que evolucionan las estrategias organizacionales. Una organización estar en nivel de madurez 5, cuando adem s de monitorizar sus estrategias de datos (nivel 4- todos los procesos de nivel 4 con una calificación de *completamente satisfecho*), obtiene un nivel de al menos *ampliamente satisfecho* para los

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

siguientes procesos del modelo de referencia de procesos orientadas a actualizar las estrategias de datos, no sólo para mejorar defectos conocidos, sino que podría ser utilizada para mejorar el rendimiento global:

- DQM.4. Mejora de Calidad de Datos.
- DG.8. Gestión de cambios en las estrategias de datos.

En la Tabla 3.9. se resumen de los niveles de madurez de MAMD.

Nivel de Madurez	Procesos
2	DM.1. Gestión de Requisitos de Datos DM.2. Gestión de la Infraestructura Tecnológica DM.3. Gestión de Datos Históricos DM.4. Gestión de la Seguridad de Datos DM.5. Gestión de la Configuración de Datos DQM.2. Monitorización y Control de Calidad de Datos DG.4. Establecimiento de Estándares, Políticas y Buenas Prácticas
3	DM.6. Gestión de Datos Maestros y de Referencia DM.7. Arquitectura y Diseño de Datos DM.8. Establecimiento de Fuentes y Destinos de Datos DM.9. Integración de Datos DQM.1. Planificación de Calidad de Datos DG.1. Establecimiento de Estrategias de Datos DG.2. Gestión del Ciclo de Vida de los Datos DG.5. Gestión de Recursos Humanos
4	DQM.3. Aseguramiento de Calidad de datos DG.3. Gestión del Valor de los Datos DG.6. Gestión de Recursos Financieros DG.7. Monitorización de las Estrategias Organizacionales de Datos
5	DQM.4. Mejora de Calidad de Datos DG.8. Gestión de Cambios en las Estrategias de Datos

Tabla 3.9. Modelo de madurez de MAMD 2.

3.8.5 Modelo de mejora

El modelo de mejora del nivel de madurez organizacional de los procesos de datos es el tercer componente que forma parte del marco de trabajo de MAMD, junto con el modelo de referencia de procesos y al modelo de evaluación. Este modelo proporciona los mecanismos necesarios para guiar la mejora del nivel de madurez organizacional de los procesos de datos y est compuesta por ocho fases (v ase Figura 3.12). Para cada una de las fases que componen la metodología, se identifican los artefactos de entrada, el procesamiento, y los artefactos de salida.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

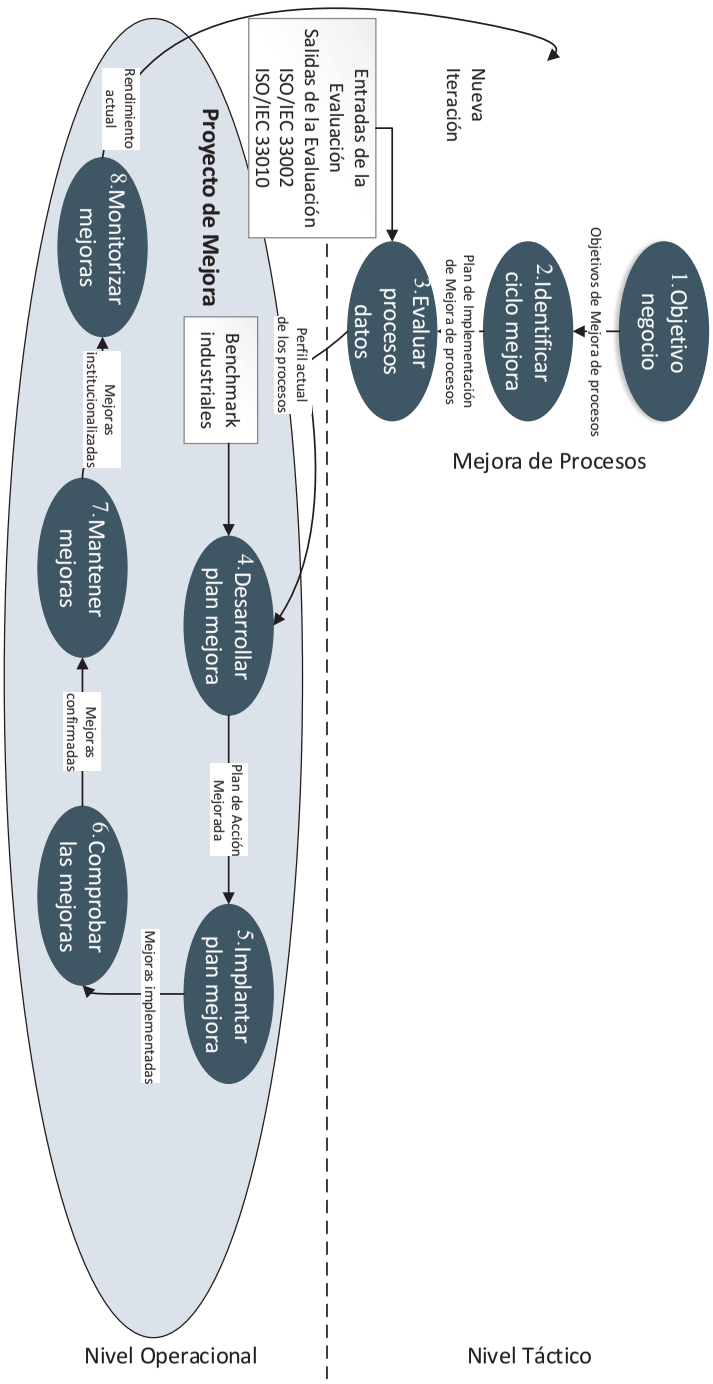

Figura 3.12. Modelo de mejora.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

Fase 1. Examinar los objetivos de negocio de la organización

El propósito de esta fase es determinar el objetivo de negocio, es decir la meta que se quiere alcanzar con la mejora. Hay que hacer hincapi en que, a trav s de la mejora de los procesos de datos, se mejora la calidad de los datos que utilizan los procesos de negocio, y que estos procesos de negocio son los que permiten cumplir el objetivo de negocio (Figura 3.13).

Figura 3.13. Repercusiones de la mejora de los procesos de datos en las organizaciones.

Estudiando la documentación proporcionada por la organización, se preparan distintas entrevistas con los empleados de la organización (a distintos niveles) identificar tanto los objetivos de negocio, así como los procesos de negocio que soportan estos objetivos. Como resultado de esta fase se obtiene un documento que explica la correcta identificación del objetivo de negocio, y los procesos de negocio que afectan a la consecución del objetivo de negocio.

Fase 2. Identificar el ciclo de mejora de procesos

El objetivo principal de esta segunda fase es determinar si la organización está capacitada para abordar un proyecto de mejora. Considerando la definición del objetivo de negocio, la documentación proporcionada por la organización, y las distintas reuniones a llevar a cabo en la organización. Para identificar el ciclo de mejora de procesos, es necesario llevar a cabo distintas reuniones con el personal de la organización, principalmente para identificar y establecer una serie de factores. Para conseguir los objetivos de esta fase, es preciso desarrollar las siguientes subfases:

Fase.2.1. Establecer la necesidad y la capacidad de mejora en la organización

En primer lugar, es necesario establecer la necesidad de mejora en la organización, así como la capacidad que tiene la organización para llevar a cabo proyectos de mejora. Por un lado, esta necesidad se establece en función de la criticidad de los problemas en la organización. Por ejemplo, en una organización que maneja información de car cter personal, la necesidad de mejora ser

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

mayor si no cumple la legislación vigente en cuanto a tratamiento de este tipo de información, que si su modelo de datos no almacena el código postal de una persona. Por otro lado, conocer la capacidad que tiene la organización para mejorar ser un aspecto clave para el enfocar el desarrollo del plan de mejora.

Para llevar a cabo ambas tareas se llevar n a cabo entrevistas con distintos miembros de la organización en las cuales se pueden utilizar los formularios propuestos en Pries-Heje y Johansen (2013).

Fase.2.2. Identificar los recursos disponibles en la organización.

En segundo lugar, ser necesario conocer los recursos de los que dispone la organización para llevar a cabo las mejoras necesarias sobre los procesos de datos. Para llevar a cabo esta tarea, se identifican y clasifican los recursos disponibles para desarrollar el plan de mejora de procesos. Los recursos disponibles se identifican y ordenan según la clasificación presentada en los catalizadores de COBIT (ISACA, 2012). En COBIT se identifican tres grandes bloques distintos de recursos:

- Información.
- Servicios, infraestructuras y aplicaciones.
- Personas, habilidades y competencias.

Fase.2.3. Identificación de roles y responsabilidades.

Por último, se identifican los distintos roles que deben existir en la organización para la implementación de las mejoras (v ase Tabla 3.10). Esta clasificación de roles viene presentada en ISO/IEC 33014.

	Organización	**Proyecto**	**Individual**
Proveedor	Gestor/patrocinador	Organización	Experto
Actuador	Unidad de mejora de procesos	Gestor del equipo de mejora	Miembro del equipo de mejora
Usuario	Cliente	Gestor del proyecto	Ejecutor

Tabla 3.10. Roles y niveles identificados para la implantación de la mejora.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

Como artefactos de salida tras la ejecución de esta fase se establece la necesidad de mejora de la organización, la capacidad para mejorar, así como los recursos disponibles en la organización para ejecutar en el futuro un plan de mejora, y los roles y responsabilidades.

Fase 3. Ejecutar la evaluación de procesos

El objetivo principal de esta fase es llevar a cabo una evaluación de los procesos de datos con el fin de conocer el grado de institucionalización de los procesos de datos en la organización, tal y como hemos visto en los apartados anteriores. Como resultado de la ejecución de esta fase se obtiene el informe de evaluación para los procesos de negocio de la organización.

Fase 4. Desarrollar un plan de mejora

El objetivo principal de esta fase es generar el plan de mejora a implementar por la organización, teniendo como objetivo principal de la mejora alcanzar o conseguir los resultados esperados para los atributos de proceso definidos en ISO/IEC 33020 que definen el nivel de capacidad, para cada uno de los procesos que sea necesario mejorar.

Teniendo en cuenta la necesidad de mejora de la organización, la capacidad para mejorar, así como los recursos disponibles en la organización para ejecutar en el futuro un plan de mejora, los roles y responsabilidades, y el informe de evaluación.

El plan de mejora tiene como objetivo adaptar los recursos organizacionales necesarios para asegurar el cumplimiento de los distintos resultados para los atributos de procesos necesarios para cada uno de los procesos de datos, de modo que se pueda alcanzar el nivel de capacidad deseado. Para desarrollar este plan de mejora es necesario ejecutar una serie de sub-fases.

Fase.4.1. Identificación de los problemas en los procesos de datos

El primer paso para desarrollar un plan de mejora es identificar los problemas en los procesos de datos que no permiten la consecución de un nivel de madurez (no conformidades). Al identificar estos problemas en los procesos de datos, se pueden destinar recursos para solventar el problema, y de este modo corregir las no conformidades para conseguir los resultados de atributo de proceso

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

necesarios para la consecución del nivel de madurez objetivo. Para identificar los problemas, se observan las evidencias recogidas durante la evaluación. A trav s de la observación de estas evidencias, se determinan cu les son esas no conformidades.

Fase.4.2. Selección de la alternativa viable de mejora que puede llevarse a cabo.

El segundo paso a la hora de llevar a cabo una mejora es la elección de la alternativa viable. El desarrollo de un plan de mejora de los procesos de datos para la consecución de un nivel de madurez objetivo puede estar enfocado de dos maneras distintas, en función de los procesos a mejorar:

- Los procesos para un determinado nivel de madurez organizacional ya existen en la organización, pero no se consigue la calificación necesaria en alguno de los atributos de proceso que definen el nivel de capacidad. En este caso, la mejora solo afectar en la dimensión de capacidad. Dado que los procesos ya est n ejecut ndose en la organización, la manera de abordar la mejora ser la de conseguir los resultados necesarios para los atributos de proceso necesarios.

- La organización est en un determinado nivel de madurez organizacional, y quiere abordar el siguiente nivel de madurez, por lo tanto, no solo tendr que aumentar el nivel de capacidad de los procesos que ya tiene implantados, sino que tendr que implantar nuevos procesos. Adem s, al igual que en el caso anterior, es necesario conseguir los resultados esperados para los atributos de proceso, para el nivel de capacidad objetivo.

Fase.4.3. Identificación de dependencias entre procesos.

Una vez se haya determinado la alternativa de mejora a llevar a cabo es necesario establecer el orden en el que se mejoran los procesos. Para ello, habr que tener en cuenta las distintas dependencias existentes entre los procesos del modelo de referencia de procesos.

Fase.4.4. Asignación de recursos disponibles.

Una vez se han identificado los problemas existentes en los procesos de datos, la alternativa y el orden de los procesos de datos, es necesario asignar recursos a cada uno de los procesos para corregir las debilidades encontradas.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

Fase.4.5. Desarrollo del plan de mejora

Una vez se han ejecutado todas las fases anteriores, se genera el plan de mejora o plan de acción para la mejora del nivel de madurez organizacional de los procesos de datos.

Fase 5. Implementar las mejoras

El objetivo principal de esta fase es la ejecución del plan definido para la mejora de los procesos de datos de la organización, para lo que es necesario en primer lugar establecer una estrategia de implantación. Una vez establecida la estrategia, es necesario para cada una de las tareas de mejora presentadas en el plan, detallar su implantación, es decir indicar de manera interna en la organización quien va a ser el responsable de realizar cada una de las tareas que componen el plan de mejora. Por último, y de manera transversal a las tareas presentadas en esta fase, ser necesario llevar a cabo una monitorización del estado de la implantación de cada una de las mejoras mientras se est n llevando a cabo.

Los artefactos de salida generados tras la ejecución de esta fase es la obtención de un nuevo nivel de capacidad y/o nivel de madurez organizacional de los procesos de datos.

Fase 6. Confirmar las mejoras

El objetivo principal de esta fase es determinar si tras la implantación del plan se ha conseguido el objetivo de la mejora, es decir se ha alcanzado el nivel de madurez objetivo. Siguiendo un m todo parecido al de la fase 3, se lleva a cabo una evaluación basada en ISO/IEC 33002 (ISO/IEC, 2015b), siguiendo el ciclo propuesto en la fase 3 para confirmar que se ha conseguido el objetivo de la mejora, es decir, que el nivel de madurez organizacional objetivo se ha conseguido. Tras la ejecución de esta fase, se obtiene como artefacto de salida un nuevo informe de evaluación de los procesos de datos.

Fase 7. Mantener las mejoras

Una vez confirmadas las mejoras, el objetivo principal de esta fase es comunicar y crear cultura a lo largo de la organización de los cambios llevados a cabo, para lo que ser necesario ejecutar una serie de sub-fases:

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

Fase.7.1 Identificar las distintas unidades y áreas de negocio afectados por la mejora

Como primer paso para mantener las mejoras en la organización, es necesario identificar los distintos departamentos afectados por la mejora. Para poder identificarlos, se observarán las áreas de la organización donde han afectado las mejoras en los procesos de datos.

Fase.7.2 Establecer los mecanismos para comunicar los cambios en los procesos y los beneficios obtenidos a toda la organización.

Una vez identificadas las unidades afectadas por la mejora, es necesario comunicar los cambios realizados para que la organización actúe conforme a la nueva manera. Esta comunicación puede ser a través de charlas informativas, por mail, carta, tel fono, etc.

Fase.7.3 Estudiar y establecer la formación necesaria a llevar a cabo para mantener las mejoras a lo largo de la organización.

Por último, una vez comunicados los cambios a los *stakeholders*, es necesario dar la formación necesaria para que se adapten a los cambios implantados. A posteriori y de forma complementaria a esta fase ser necesario la realización de evaluaciones para confirmar que las mejoras han sido mantenidas, y como estos cursos de formación que inicialmente fueron adoptados por la mejora de un conjunto especifico de procesos, afectan a otras áreas de negocio.

Fase 8. Monitorizar el rendimiento de los procesos de negocio

En esta última fase, el objetivo es comprobar como la mejora llevada a cabo sobre los procesos de datos de la organización se ve reflejada en el proceso de negocio y en el objetivo de la organización. Para ello, tendremos en cuenta los indicadores clave de rendimiento (KPI-*Key Performance Indicator*) definidos en la organización tanto para los procesos de datos y negocio, así como para el objetivo de la organización. Se obtendrá una serie de conclusiones sobre el grado en el que la mejora en los procesos de datos ha afectado al desempeño del proceso de negocio para el objetivo de negocio que se perseguía.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

3.8.6 Comparación entre los modelos de referencia de procesos

En la Tabla 3.11 se resume una comparación entre el modelo de referencias de procesos de MAMD, DMM e ISO 8000-61. Como se puede ver ISO 8000-61 se centra en los procesos de gestión de calidad de datos, no tiene procesos específicos de gobierno, aunque algunos de los procesos sí que estén relacionados con los procesos de gobierno de MAMD.

En cuanto al modelo DMM se puede decir que ste sí que tiene procesos específicos de gobierno, gestión y gestión de calidad de datos. Por último, relativo a los modelos de madurez tanto MAMD como el de ISO 8000-62 (asociado a ISO 8000-61) basan sus modelos de madurez en el est ndar ISO/IEC 33000 mientras que DMM se basa en el modelo de madurez de CMMI.

MAMD	ISO 8000-61	DMM
DM.1. Gestión de Requisitos de Datos	DQP.1.	DO 4.1
DM.2. Gestión de la Infraestructura Tecnológica	DRS.2.	PA 5.1, PA 5.3
DM.3. Gestión de Datos Históricos		PA 5.5
DM.4. Gestión de la Seguridad de Datos	DRS.4.	PA 5.3
DM.5. Gestión de la Configuración de Datos		SP 6.5
DM.6. Gestión de Datos Maestros y de Referencia		DMS 1.1, DG 2.1
DM.7. Arquitectura y Diseño de Datos		DG 2.2, DG 2.3
DM.8. Establecimiento de Fuentes y Destinos de Datos	DRS.2.	DO 4.3
DM.9. Integración de Datos	DQA.2.	PA 5.4
DQM.1. Planificación de Calidad de Datos	DQP.4	SP 6.1, DQ 3.2,
DQM.2. Monitorización y Control de Calidad de Datos	DQC.3	SP 6.1
DQM.3. Aseguramiento de Calidad de datos	DQA.1, DQA.2, DQA.3, DQA.4	DQ 3.3
DQM.4. Mejora de Calidad de Datos	DQI.1, DQI.2, DQI.3	DQ 3.4
DG.1. Establecimiento de Estrategias de Datos	DQP.2., DQC.1	DMS 1.1, DQ 3.1, DG 2.1, DMS 1.2
DG.2. Gestión del Ciclo de Vida de los Datos		DO 4.2
DG.3. Gestión del Valor de los Datos	DQP.3., DQA.2.	
DG.4. Establecimiento de Estándares, Políticas y Buenas Pr cticas	RPV.2.	DG 2.1, DMS 1.2, DQ 3.1, DQ 3.2, PA 5.2
DG.5. Gestión de Recursos Humanos		PA 5.2
DG.6. Gestión de Recursos Financieros	DQC.3.	DMS 1.5
DG.7. Monitorización de las Estrategias Organizacionales de Datos		DG 2.1, DMS 1.1
DG.8. Gestión de Cambios en las Estrategias de Datos		DMS 1.1

Tabla 3.11. Comparación entre MAMD, ISO 8000-61 y DMM.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

La comparación de MAMD con DAMA es m s compleja, se puede observar que DAMA tiene procesos que específicamente tratan la gestión de los metadatos, mientras que en el caso de MAMD se les presta mayor atención a los datos maestros. En la Tabla 3.12 se presenta el resultado la comparación entre MAMD y DAMA.

DAMA	MAMD
Gobierno de datos	Los ocho (8) procesos de gobierno de datos (DG.1 – DG.8)
Arquitectura de datos	DM.7. Arquitectura y diseño de datos
Diseño y modelado de datos	DM.1, DM.2, DM.6, DM.7
Operaciones y almacenamiento de datos	Varios procesos de gestión de datos (DM)
Seguridad de datos	DM.4. Gestión de seguridad de datos
Integración e interoperabilidad de datos	DM.8, DM.9
Gestión de documentación y contenido	DM.1, DM.5, DM.7, DG.4
Datos maestros y de referencia	DM.6
Inteligencia de negocio y data warehousing	DM.2, DG.1, DG.3
Metadatos	
Calidad de datos	DQM.1, DQM.2, DQM.3, DQM.4

Tabla 3.12. Estudio comparativo de los procesos de MAMD y DAMA

3.8.7 Ejemplos de utilización de MAMD

Esta sección resume dos casos prácticos en los que se ha utilizado MAMD.

3.8.7.1 HOSPITAL

Este estudio de caso se desarrolló en el departamento de informática de una facultad de medicina (llamado CSD por motivos confidenciales). El CSD se encarga de proporcionar servicios inform ticos a los seis (6) departamentos acad micos del hospital universitario. El número de empleados de veintidós (22) personas, que dan soporte a la mayoría de los procesos de negocio de los diferentes departamentos.

El alcance de la evaluación se decidió en una reunión con los responsables del CSD y el resultado fue realizar una evaluación de nivel 2. De acuerdo con la identificación de los procesos de negocio considerados como parte de la

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

evaluación y teniendo en cuenta el inter s del CSD en optimizar los procesos del departamento más críticos se eligieron los siguientes:

- **Proceso Principal (PP):** Mantenimiento de repositorios de datos farmacológicos. Este proceso de negocio se encarga de recuperar, documentar y unificar los resultados obtenidos en diferentes bancos de pruebas farmacológicos hechos en diferentes experimentos.

- **Proceso Auxiliar 1 (PA1):** Generación de informes bioestadísticos. Este proceso se encarga de soportar la generación de varios reportes requeridos por otros equipos del CSD, principalmente de terceros ajenos al departamento.

- **Proceso Auxiliar 2 (PA2):** Mantenimiento del software clínico. Este proceso es ejecutado por el equipo de inform tica.

Para cada proceso del nivel de madurez 2 se obtuvieron las siguientes conclusiones después de la inspección de las correspondientes evidencias:

- **DM.1. Gestión de requisitos de datos.** El PP est totalmente descrito en la documentación interna (de hecho, el CSD sigue la norma ISO 9001) y consecuentemente, el equipo farmacológico tiene identificados apropiadamente los requisitos de datos durante la descripción del proceso. Con respecto a la gestión de los procesos, los objetivos en la gestión de requisitos estaban identificados, y el proceso est planeado. Con respecto a la gestión de los productos de trabajo, hay plantillas predefinidas para la recolección de requisitos de datos y varios controles y revisiones de los documentos generados realizadas. Así que AP.2.1 y AP.2.2 pueden ser considerados como ampliamente satisfechos.

- **DM.2. Gestión de la infraestructura tecnológica.** El proceso principal est convenientemente soportado por la infraestructura tecnológica interna del CSD. La gestión de la infraestructura tecnológica est a cargo de los miembros del equipo de inform tica, quien es responsable de recoger los requisitos específicos, validar y desarrollar las soluciones tecnológicas que soportan los trabajos de los equipos. En este sentido, tanto AP.2.1 como AP.2.2 est n *totalmente satisfechos*.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

- **DM.3. Gestión de datos históricos.** Se obtuvieron suficientes evidencias para manifestar que el equipo farmacológico mantiene estrictas políticas sobre la gestión de datos históricos porque es importante para el xito de los reportes y est n soportados por el equipo de inform tica. Esto hace que el AP.1.1 se considere como *totalmente satisfecho*. Como este proceso es muy importante para el equipo farmacológico ponen especial atención a la gestión, así como a la gestión de productos de trabajo. Por lo tanto, ambos AP.2.1 y AP.2.2 pueden ser considerados como *totalmente satisfechos*.

- **DM.4. Gestión de seguridad de datos.** El equipo farmacológico especificó los requisitos de seguridad de datos, que también incluían la legislación de privacidad y reglas del país del CSD. Esto *satisface totalmente* AP.1.1 para hacer operativas las restricciones de seguridad de los datos, el equipo de inform tica desarrolló los correspondientes mecanismos para soportar el proceso principal, *satisfaciendo totalmente* tanto AP.2.1 como AP.2.2.

- **DM.5. Gestión de la configuración.** Los cambios hechos sobre la definición de los datos a lo largo de los años cuando se ejecuta el proceso principal no han estado adecuadamente gestionados. Por lo tanto, el equipo farmacológico no ha podido aportar ninguna evidencia que nos permitiese llegar a la conclusión de que el AP.1.1 está satisfecho. Observando el proceso auxiliar 2, se pudo verificar que sí llevaban a cabo la gestión de configuración del software, y sí que aplicaban estas tareas a los datos. Por tanto, de nuevo se obtuvo la conclusión final sobre la calificación de este proceso de gestión de la configuración, el AP.1.1 se satisface, pero sin embargo no se satisface ni AP.2.1, ni AP.2.2.

- **DQM.2. Monitorización y Control de Calidad de Datos.** Como han estado trabajando tiempo atrás, el equipo farmacológico ha establecido ciertas acciones respecto a la calidad de datos que usan para comprobar la validez y consistencia de los datos que reciben. Desde las respuestas que dieron en los cuestionarios y parte de los trabajos que mostraron, se pudo verificar que ejecutan ciertas acciones que podían considerarse como acciones de monitorización de calidad de datos. Consecuentemente, concluimos que el AP.1.1 se puede calificar como *totalmente satisfecho*. Sin embargo, incluso cuando han demostrado suficiente conocimiento para llevar a cabo tareas de monitorización, no siguen un proceso definido, y no pueden dar, por

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

tanto, evidencia de ello. Por lo tanto, AP.2.1 y AP.2.2 se consideran como *no satisfechos*. Al inspeccionar el proceso auxiliar 1, se descubrió que el equipo de bioestadística seguía un proceso definido, con los recursos necesarios y las responsabilidades claramente designadas. Además, tenían documentados todos aquellos aspectos que conciernen a los resultados de las monitorización y control de calidad de datos. Durante las entrevistas, el equipo de bioestadística dio respuesta y evidencias que permitieron establecer no sólo el AP.1.1 totalmente satisfecho, sino tambi n el AP.2.1 y el AP.2.2 como *ampliamente satisfechos*.

- **DG.4. Definición de estándares, políticas y buenas prácticas.** Como parte de la gestión del equipo de farmacología, ha adaptado e implementado una serie de políticas, pr cticas y procedimientos que están basados en las políticas generales del CSD. Muchas de estas políticas est n orientadas a la gestión de los datos y la estimación de su calidad. Mientras se realizaban las preguntas de los cuestionarios no solo respondieron, sino que aportaron otras pruebas que nos permitieron situar los resultados de este proceso como realizado (y consecuentemente el AP.1.1 est *totalmente satisfecho*). El trabajo de definición de un catálogo de estándares, políticas y procedimientos no se ejecuta de manera predefinida, no hay responsabilidades asignadas y los recursos se proporcionan sólo cuando son necesarios. Así que, no se pudo obtener ninguna evidencia que fuese suficiente para calificar el AP.2.1 y AP.2.2 ni si quiera como satisfechos.

Bas ndose en las conclusiones que se obtuvieron a partir del an lisis de las evidencias se obtuvieron, los resultados que se resumen y muestran en la Tabla 3.14 para cada uno de los procesos y la respectiva calificación de sus atributos de proceso (mostrada como: N, P, L, F).

AP.2.2	L	F	F	F	N	L	F	F
AP.2.1	L	F	F	F	N	L	F	F
AP.1.1	F	F	F	F	F	F	F	F
	DM.1	**DM.2**	**DM.3**	**DM.4**	**DM.5**	**DQM.2**	**DG.2**	**DG.4**
	Nivel de Madurez 2							

Tabla 3.13. Resultados obtenidos en cada proceso en la evaluación MAMD del CSD.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

En este estudio de caso, como se puede ver en la Tabla 3.13, todos los procesos han obtenido una calificación de "F" para el AP.1.1. Esto permite afirmar que la organización tendría consolidado el nivel de madurez 1. A pesar de esto, y dado que DM.5. ha obtenido un valor "N" para los AP.2.1 y AP.2.2, no se consolida En esta situación, el nivel de madurez conseguido m s alto es el 1 y consecuentemente se considera que CSD est en el nivel de madurez 1.

Un posible plan de mejora para consolidar el nivel 2, debería orientar los esfuerzos a conseguir que el proceso DM.5. pasará de "N" a "L".

3.8.7.2 BIBLIOTECA UNIVERSITARIA DE LA UNIVERSIDAD DE CASTILLA-LA MANCHA (UCLM)

En este estudio de caso, la aplicación del marco de trabajo MAMD ha sido ala Biblioteca Universitaria de la UCLM, que da soporte a la comunidad universitaria formada por 31.203 estudiantes, 1.044 personas de administración y servicios y 2.227 miembros de personal acad mico y de investigación (censo de 2014) situados en diferentes campus. Tiene un total de 13 puntos de servicio y 3 salas de lectura con una ratio de 6,6 estudiantes por plaza de estudio. En 2014, la Biblioteca Universitaria gestiona una colección bibliográfica de 1.140.509 volúmenes, así como acceso a 46.411 revistas electrónicas y 487.077 libros electrónicos, aparte de otros servicios. Desde estos recursos, se cuenta un total de 289.388 préstamos de material bibliográfico, un total de 339.279 descargas de libros electrónicos y un total de 588.520 búsquedas en base de datos de referencia. A todo esto, le da soporte un total de 78 profesionales de diferentes categorías y rangos para todas las utilidades que prestan.

Además de la gestión de la literatura, la Biblioteca Universitaria tambi n gestiona otros servicios como pr stamo de equipamiento inform tico tanto para el aprendizaje como para la enseñanza, cursos, gestión de redes sociales, etc. Toda esta actividad se soporta por una infraestructura tecnológica gestionada por el centro de datos de la universidad, que permite el uso no sólo del equipamiento que utiliza el personal, sino tambi n de 964 puestos en las salas de lectura. Tanto los recursos humanos como los tecnológicos se usan en la integración de datos con otras instituciones, algo interesante y a tener en cuenta para este trabajo.

Se decidió evaluar hasta el nivel 2 del modelo de madurez de MAMD, al igual que en el estudio de caso del CSD. Se eligieron los siguientes procesos:

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

- **Proceso principal (PP):**

para la recolección de requisitos de datos y se realizan varios controles y revisiones de los documentos generados. Adem s, la organización tenía un repositorio donde se especificaban los requisitos de datos.

- **DM.2. Gestión de la infraestructura tecnológica.** El PP est soportado adecuadamente por una infraestructura tecnológica integrada completamente en los servicios TIC de la UCLM. La gestión de la infraestructura propiamente dicha se realiza por tanto de forma externalizada, pero en la Biblioteca Universitaria existen responsables t cnicos que se encargan de recopilar, validar e informar de los cambios que hay que hacer en la infraestructura tecnológica requisitos al área TIC. Los responsables del rea TIC mostraron la infraestructura e incluso se observaron los sistemas de monitorización que se encargan, por ejemplo, de alertar en caso de que hubiese problemas de disponibilidad.

- **DM.3. Gestión de datos históricos.** La Biblioteca Universitaria mantiene una serie de políticas muy detalladas, entre las que se encuentran las correspondientes a la gestión de datos históricos. En el caso del PP, descatalogar un libro significa mover sus datos del fondo documental en producción al fondo documental descatalogado y gestionar las consecuencias de que un documento pase a estar descatalogado (por ejemplo, cancelar los pr stamos asignados a los usuarios). Este proceso es realizado por el responsable del fondo documental, que es el único que tiene autoridad para descatalogar una obra. En el caso, del PA1, se tiene como política borrar los datos correspondientes a la circulación de fondos que tenga m s de tres años de antigüedad. Con respecto al PA2, se tiene como política borrar los datos de los usuarios al comienzo del curso siguiente al que terminaron sus estudios. Sí que pudimos constatar que la ejecución de la gestión de datos históricos en cualquiera de las instancias de los procesos est completamente automatizada, con lo que los productos de trabajo est n adecuadamente gestionados. Adem s, cabe destacar un proceso específico de desafección de fondos que se encuentra perfectamente definido y que trata los datos históricos de los fondos de la biblioteca.

- **DM.4. Gestión de seguridad de datos.** La Biblioteca Universitaria es muy cuidadosa en lo que se refiere a los datos personales. De hecho, hereda de la universidad la totalidad de los procedimientos de protección de los datos. Esto se traduce en que, para cualquiera de las tres instancias de los procesos observados, se gestiona adecuadamente

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

el acceso a los datos, impidiendo que usuarios no autorizados consulten o modifiquen datos. No obstante, los datos más sensibles se mueven en el proceso PA1, ya que son datos de usuarios y las obras que le son prestadas, y que pueden revelar información confidencial acerca de aspectos sensibles de los usuarios de la biblioteca. En el otro extremo estaban los datos de las obras, cuya gestión no requería especial protección. Precisamente por eso, se decidió inspeccionar la instancia de PA1. Se concluyó que la ejecución de las políticas correspondiente a la gestión de datos históricos efectivamente se realiza por las personas a quienes se les ha asignado la responsabilidad de hacerlo, que es un proceso planificado que usa los recursos que le han sido asignados, aunque no se monitoriza y est perfectamente adecuado. Por otro lado, el cat logo de política de seguridad responde a un esquema madurado a lo largo de la vida de la Biblioteca Universitaria, y que por tanto se documenta, controla y transfiere de forma precisa.

- **DM.5. Gestión de la configuración.** La Biblioteca Universitaria est plenamente adaptada a la normativa existente en la red de bibliotecas de España. Ello implica tener en cuenta y propagar a los procesos de negocio de la biblioteca las correspondientes modificaciones y evoluciones en la definición de los datos. Por ejemplo, en el caso del PP, cuando se asimiló el formato MARC 21, se llevó a cabo un proceso de gestión de configuración y versionado del modelo de los datos para que las nuevas definiciones fueran coherentes en todas las aplicaciones implicadas.

- **DQM.2. Control y monitorización de calidad de datos.** Es interesante resaltar que los gestores de la Biblioteca Universitaria son muy conscientes de la importancia de tener un nivel adecuado de calidad de datos. Por ejemplo, para comunicarse con otras bibliotecas de otras universidades o para integrar datos de usuarios externos. Se observó que realizaban monitorización de calidad de datos, aunque no se aprovechaban los resultados obtenidos para llevar a cabo acciones en base a la información obtenida.

- **DG.4. Establecimiento de estándares, políticas buenas prácticas y procedimientos.** Como parte de la gestión de la Biblioteca Universitaria, se han definido e implementado adecuadamente una serie de políticas, buenas prácticas y procedimientos que están basadas en las políticas generales de la UCLM. Muchas de estas políticas están orientadas a

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

gestionar adecuadamente los datos

Cabe destacar, como último paso de este estudio de caso que, la Biblioteca Universitaria superó satisfactoriamente el proceso de auditoría llevado a cabo por AENOR y obtuvo la concesión de AENOR del certificado de nivel 2 de madurez organizacional de los procesos de datos por MAMD – ISO 8000 – ISO/IEC 33000.

AENOR

Certificado de Nivel de Madurez
Procesos de Gobierno y Gestión de la Calidad de Datos

AENOR ER Madurez Datos 2 ISO 33000

NMD-2018/0001

AENOR certifica que la organización

UNIVERSIDAD CASTILLA-LA MANCHA
BIBLIOTECA UNIVERSITARIA

se ha evaluado la madurez de los procesos de gobierno y gestión de la calidad de datos conforme con la Norma ISO/IEC 33000 e ISO/IEC 8000 (MAMD)

y se ha alcanzado el nivel: 2

para las actividades: Gestión de Datos

que se realizan en: Biblioteca Universitaria (servicios centrales). Avda. de Camilo José Cela, 6, 13071 Ciudad Real – España
Servicio de Bibliotecas del Campus de Albacete. Edificio José Prat, Plaza de la Universidad, s/n, 02071 Albacete - España
Servicio de Bibliotecas del Campus de Ciudad Real. Edificio Bernardo Balbuena. Avda. Camilo José Cela, s/n. 13071 Ciudad Real - España
Servicio de Bibliotecas del Campus de Cuenca. Avda. de los Alfares, 42, 16071 Cuenca - España
Servicio de Bibliotecas del Campus de Toledo. Edificio Madre de Dios, Cobertizo de San Pedro Mártir, s/n. 45071 Toledo - España

Fecha de primera emisión: 2018-02-26
Fecha de expiración: 2021-02-26

Rafael GARCÍA MEIRO
Director General

AENOR INTERNACIONAL, S.A.U.
Génova, 6. 28004 Madrid. España
Tel. 91 432 60 00.- www.aenor.com

Original Electrónico

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

3.9 LECTURAS RECOMENDADAS

- Van Looy, A. (2014). *Business process maturity: A comparative study on a sample of business process maturity models*. Springer Science & Business Media.

  Ofrece una guía completa para organizar y clasificar las disciplinas de gestión de procesos de negocio. Incluye un tutorial paso a paso para ayudar a los lectores a hacer una selección rápida y confiable y presenta una descripción general de 69 modelos de madurez de procesos de negocios, que complementan a los presentados en este capítulo.

3.10 SITIOS WEB

- *http://iso8000.es/*

  Este portal sirve de foro para reunir información relativa a los procesos de gestión de calidad de datos (ISO 8000-6x), y con respecto a la gestión de datos maestros (ISO 8000-1x0).

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:49:23.

4

CALIDAD DE REPOSITORIOS DE DATOS

4.1 MODELO DE CALIDAD DE DATOS

Un modelo de calidad de datos es un conjunto de dimensiones o características de calidad de datos. Un modelo de calidad de datos permite juzgar la calidad de un determinado repositorio de datos. A lo largo de la historia de la calidad de los datos se han propuesto numerosos modelos de calidad de datos, conteniendo diversas clasificaciones y dimensiones de las dimensiones de calidad de datos desde distintos puntos de vista.

Uno de los modelos de calidad de datos más importantes es el proporcionado por la norma internacional ISO/IEC 25012, que recoge las principales características de calidad que cualquier conjunto de datos un sistema de información debe poseer. Estas características se agrupan en tres bloques diferentes:

- **Calidad de los datos inherente:** incluye aquellas características cuya medición depende de la propia naturaleza de los datos.
- **Calidad de datos dependiente del sistema:** recoge aquellas cuya medición depende del sistema de información en el que los datos están almacenados.
- **Calidad de datos inherente y dependientes del sistema:** incluye las características cuya medición depende tanto de la naturaleza de los

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-28 07:11:56.

datos como de la definición del sistema de información en el que los datos se alojan.

La Tabla 4.1 muestra la clasificación de las características de calidad de datos de ISO/IEC 25012, que se describen a continuación.

Características	Inherentes	Dependientes del Sistema
Exactitud	X	
Completitud	X	
Consistencia	X	
Credibilidad	X	
Actualidad	X	
Accesibilidad	X	X
Conformidad	X	X
Confidencialidad	X	X
Eficiencia	X	X
Precisión	X	X
Trazabilidad	X	X
Comprensibilidad	X	X
Disponibilidad		X
Portabilidad		X
Recuperabilidad		X

Tabla 4.1. Características de Calidad de Datos de ISO/IEC 25012

A continuación, se introducen las definiciones de las dimensiones de calidad de datos (ordenadas por orden alfabético) de acuerdo a ISO 25012:

- **Accesibilidad:** *"grado en que los datos pueden ser accedidos, en un contexto específico de uso, particularmente por personas que necesitan una tecnología de soporte o una configuración especial debido a alguna incapacidad o minusvalía"*.

- **Actualidad:** *"grado en que los datos tienen atributos que son de la edad adecuada en un contexto específico de uso"*, en definitiva, se

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-28 07:11:56.

puede interpretar como *"la medida en que los datos se actualizan correctamente y en el momento oportuno"*.

- **Completitud:** *"grado en que los datos de sujetos asociados con una entidad tienen valores para todos los atributos esperados e instancias de entidades relacionadas en un contexto específico de uso"*. Definida también como *"medida en que los datos asociados a una entidad tienen valores para todos los atributos necesarios para representar a la entidad"*.

- **Comprensibilidad:** *"grado en el que los datos tienen atributos que permiten su lectura e interpretación por los usuarios, y están expresados en lenguajes, símbolos y unidades apropiados en un contexto específico de uso"*.

- **Confidencialidad:** *"grado en el que los datos tienen atributos que aseguran que sólo pueden ser accedidos por usuarios autorizados en un contexto específico de uso"*.

- **Conformidad:** *"grado en el que los datos cumplen estándares, convenciones o legislaciones o reglas similares relativas a la calidad en un contexto específico de uso"*.

- **Consistencia:** *"grado en que los datos tienen atributos que están libres de contradicción y son coherentes con otros datos en un contexto específico de uso"*.

- **Credibilidad:** *"grado en que los datos tienen atributos que son considerados como verdaderos y creíbles por los usuarios en un contexto específico de uso"*.

- **Disponibilidad:** *"grado en el que los datos pueden ser recuperados por usuarios y/o aplicaciones autorizados en un contexto específico de uso"*.

- **Eficiencia:** *"grado en el que los datos pueden ser procesados y proporcionan los niveles esperados de rendimiento utilizando cantidades y tipos apropiados de recursos en un contexto específico de uso"*.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-28 07:11:56.

- **Exactitud:** *"medida en que los datos tienen atributos que representan correctamente el verdadero valor del atributo de un concepto o evento en un contexto específico de uso"*. También se define como *"proximidad del valor de datos a un conjunto de valores válidos predefinidos en un dominio dado para un contexto específico"*.
- **Portabilidad:** *"grado en el que los datos pueden ser instalados, reemplazados o trasladados desde un sistema a otro sin perder la calidad existente en un contexto específico de uso"*.
- **Precisión:** *"grado en el que los datos son exactos o proporcionan discriminación en un contexto específico de uso"*.
- **Recuperabilidad:** *"grado en el que los datos tienen atributos que permiten mantener y preservar un nivel especificado de operabilidad y calidad, incluso en caso de fallos, en un contexto específico de uso"*.
- **Trazabilidad:** *"grado en el que los datos proporcionan un camino de auditoría de acceso a los datos y sobre los cambios realizados en los datos en un contexto específico de uso"*.

Estas características de calidad de los datos se cuantifican mediante varias medidas que se resumen a continuación.

4.2 MEDIDAS DE CALIDAD DE DATOS

La norma ISO/IEC 25024 (ISO/IEC, 2015a) propone un conjunto de medidas con el objetivo de cuantificar el grado en el que un conjunto de datos cumple las características del modelo de calidad. El valor de la calidad de esas características se calcula a través de un conjunto de propiedades de calidad de datos. Cada característica o propiedad del Modelo de Calidad de Datos toma un valor dentro del intervalo [0, 100]. El valor de las propiedades de calidad de datos es una combinación de los valores de un conjunto de medidas de calidad de datos que están basadas en los métodos de medición de ISO/IEC 25024.

A modo de ejemplo, presentamos a continuación algunas propiedades y sus medidas de calidad:

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-28 07:12:09.

Exactitud sintáctica de los datos

La exactitud sintáctica de los datos mide la relación de atributos que tienen valores relacionados sintácticamente precisos. Por ejemplo, un correo electrónico debe cumplir una regla sintáctica que especifique lo que se considera un correo electrónico válido, por ejemplo: "user@organization.com".

La medida asociada es:

Exactitud Sintáctica de datos	Cercanía de los valores de datos a un conjunto de valores definido en un dominio	**X=A/B** A=número de elementos de datos con valores sintácticamente exactos B=número de elementos de datos considerados	Aplicable durante todas las etapas del ciclo de vida de los datos Elementos de datos: Atributos de un registro Registros de un fichero

Completitud de registro

La completitud de registro mide el ratio de atributos con un valor asociado no nulo en un registro. Es importante resaltar que los atributos necesarios para un registro no tienen por qué limitarse a los atributos que lo identifican (estos serían un subconjunto de los necesarios), ya que los atributos necesarios incluyen valores importantes para aplicaciones y procesos de negocio. Por ejemplo, para identificar un cliente es necesario un número de cliente, un correo electrónico, el primer y segundo nombre y un número de cuenta bancaria.

Completitud de Registro	Completitud de los elementos de datos de un registro en un fichero	**X=A/B** A=número de elementos de datos con valores no vacíos B=número de elementos de datos considerados	Aplicable durante todas las etapas del ciclo de vida de los datos Elementos de datos: Atributos de un registro Registros de un fichero

Integridad referencial

Integridad referencial mide las conexiones a través de los datos. Los datos deben conectarse correctamente, es decir, un mismo valor debe existir para atributos conectados por una relación entre ficheros. Así, Ejemplo: Los productos en un pedido deben estar en el catálogo de productos.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-28 07:12:09.

Integridad Referencial	Coherencia en los valores de atributos que enlazan distintas entidades. Por ejemplo, para cada valor de un atributo en un fichero, existe el mismo valor del mismo atributo en el fichero relacionado	**X=A/B** A=número de elementos de datos consistentes en la relación B=número de elementos de datos considerados	Aplicable durante todas las etapas del ciclo de vida de los datos Elementos de datos: Atributos de un registro Registros de un fichero

Credibilidad de la fuente

La credibilidad de la fuente mide el *"grado en que los valores son proporcionados por una organización calificada, validada y confiable"*. Esta métrica se puede interpretar como la medida en que los valores de los datos son proporcionados por una entidad confiable o han sido validados por un evaluador independiente. Por ejemplo, se considerarán confiables los valores de los datos provenientes del Gobierno.

Credibilidad de la fuente	Grado en el que los valores de datos son proporcionados por una organización cualificada	**X=A/B** A=número de valores de datos proporcionados o validados/certificados por una entidad cualificada B=número de valores de datos considerados	Aplicable durante la recolección y adquisición de datos Elementos de datos: Valores de un atributo Registros de un fichero

Actualidad

La actualidad mide el *"grado en que los elementos de datos se actualizan oportunamente"*. Así, por ejemplo: El estado del envío debe actualizarse después de cada evento (por ejemplo, orden, pago, despacho, entrega, ...).

Actualidad	Grado en el que los datos son actualizados convenientemente	**X=A/B** A=número de elementos de datos convenientemente actualizados B=número de elementos de datos considerados	Integración de datos, procesamiento y almacenamiento Elementos de datos: Atributos de un registro Registros de un fichero

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-28 07:12:09.

4.3 PROCESO DE EVALUACIÓN

Para la evaluación de la calidad de datos se puede adoptar la norma internacional ISO/IEC 25040, que propone un modelo de evaluación, que considera tanto las entradas al proceso de evaluación (requisitos para la evaluación, especificación de requisitos de calidad, productos a evaluar, etc.), como las restricciones (necesidades, planificación, etc.) y los recursos disponibles (personal, herramientas, equipos informáticos, etc.) para obtener las correspondientes salidas (plan de evaluación, medidas, criterios de decisión, resultados e informe de evaluación, etc.).

En la Figura 4.1 se resume las principales actividades de este proceso.

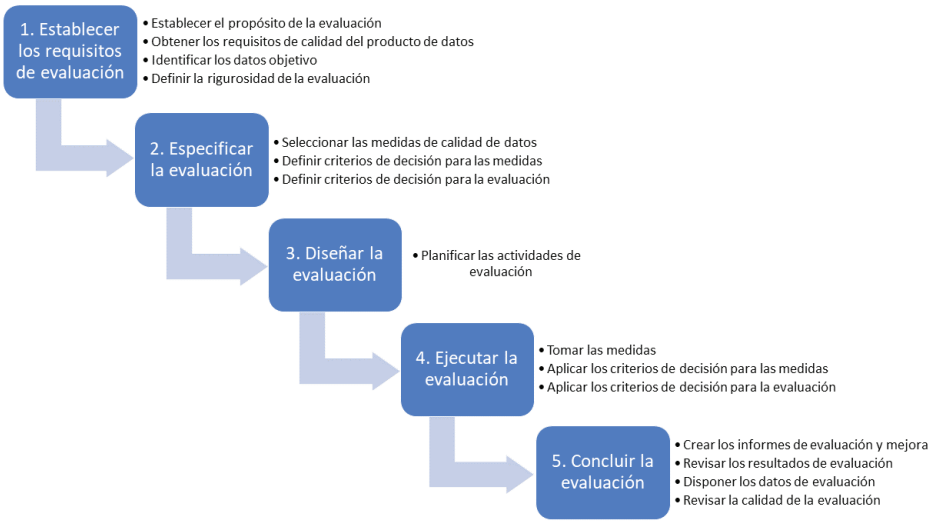

Figura 4.1. Proceso de evaluación de calidad de datos basado en ISO/IEC 25040

4.3.1 Establecer los requisitos de evaluación

Esta actividad consta de cuatro tareas:

- Establecer el propósito de la evaluación.
- Obtener los requisitos de calidad de los productos de datos.
- Identificar las partes del producto de datos a evaluar.
- Definir la severidad de la evaluación.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:28.

Algunas de estas tareas se detallan en las siguientes subsecciones. Las entradas y los resultados, los recursos necesarios y los roles involucrados se muestran en la Figura 4.2.

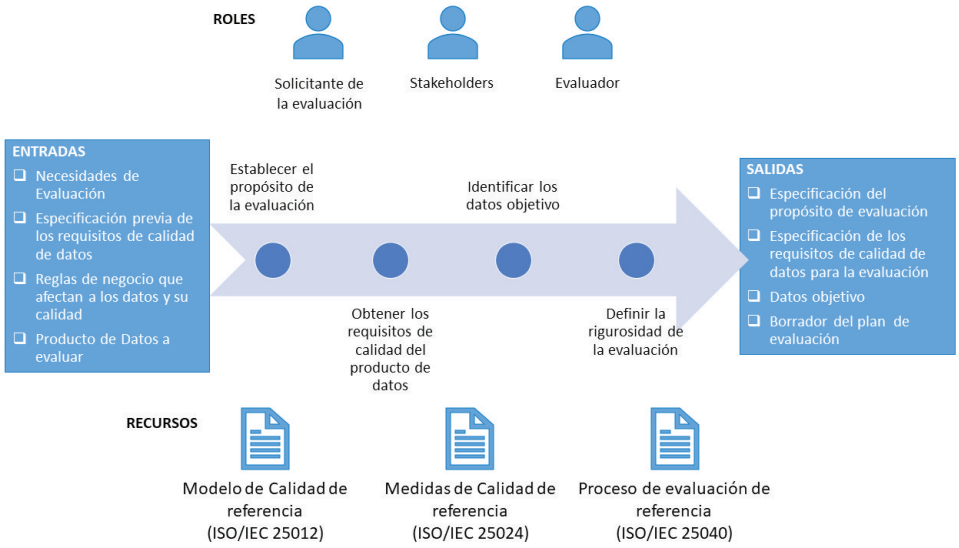

Figura 4.2. Actividad 1 – Establecer los requisitos de evaluación

4.3.1.1 ESTABLECER EL PROPÓSITO DE LA EVALUACIÓN

El propósito de la evaluación de la calidad del producto de datos se documentará como base para las actividades y tareas de evaluación adicionales, y puede ser:

- Asegurar la calidad del conjunto de datos;
- Decidir sobre la aceptación de un conjunto de datos adquiridos;
- Controlar la calidad del producto de datos;
- Detectar defectos de datos para decidir cuándo mejorar la calidad de los productos de datos;
- Comparar el producto de datos con productos competitivos;
- Seleccionar un producto de datos entre productos alternativos;
- Evaluar los efectos positivos y negativos de un producto de datos cuando se utiliza;
- Determinar la causa de los defectos de datos en una investigación.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:28.

4.3.1.2 OBTENER LOS REQUISITOS DE CALIDAD DE LOS PRODUCTOS DE DATOS

Se identificarán los *stakeholders* que conocen las reglas de negocio que deben cumplir los datos y que pueden afectar a su calidad. Los *stakeholders* pueden ser los dueños de datos (los responsables finales de datos), gobernadores de datos, arquitectos de datos, analistas de datos, administradores de datos, etc.

Una regla de negocio puede:

- Referirse explícitamente a los datos, por ejemplo:
  - *"La edad de los empleados debe ser mayor de 23 años"*.
  - *"Los correos electrónicos de los empleados terminarán con el dominio de la empresa (company.com)"*, o.
- Afectar de forma implícita a los datos, por ejemplo:
  - *"La empresa toma datos de tres proveedores de datos diferentes, pero sólo se confía en los datos de A y B y se utilizan para las principales operaciones de negocio"* – esta regla de negocio implica que los datos de los proveedores de datos A y B se consideran confiables y creíbles mientras que los datos del proveedor de datos C no son completamente confiables–.
  - *"Los horarios de los trenes se actualizan cada 5 minutos para que los pasajeros conozcan los retrasos"* – esto implica que los datos sobre los horarios del tren deben actualizarse cada 5 minutos).

Una especificación completa debe incluir la lista de reglas de negocio relacionadas con los datos objetivo, clasificadas por las diferentes características de la calidad de los datos (por ejemplolas reglas de negocio anteriores se clasificarían respectivamente en consistencia, exactitud, credibilidad y actualidad).

Si la especificación de requisitos de calidad de datos existe y está disponible puede ser reutilizada, revisada y refinada. Si no existe una especificación previa de las reglas de negocio, se deben descubrir las reglas de negocio por medio de diferentes métodos, incluyendo: restricciones de datos implementadas dentro de la base de datos, solicitudes de documentación y entrevistas con las partes interesadas.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:28.

Adicionalmente, se deben solicitar documentación adicional sobre los datos, incluyendo documentación sobre los procesos y aplicaciones que administran y controlan los datos de destino.

4.3.1.3 DEFINIR LA RIGUROSIDAD DE LA EVALUACIÓN

La rigurosidad se define con el fin de proporcionar confianza en la calidad del producto de datos de acuerdo con el uso y propósito de la evaluación. La rigurosidad de la evaluación debe estar relacionada con un conjunto de características que establecen los niveles de evaluación esperados que definen las técnicas de evaluación a aplicar.

4.3.2 Especificar la evaluación

Esta actividad, que se resume en la Figura 4.3, consta de cuatro tareas:

- Seleccionar las medidas de calidad de datos (módulos de evaluación).
- Definir los criterios de decisión para las medidas de calidad de datos.
- Establecer los criterios de decisión para la evaluación de la Calidad de los Datos.

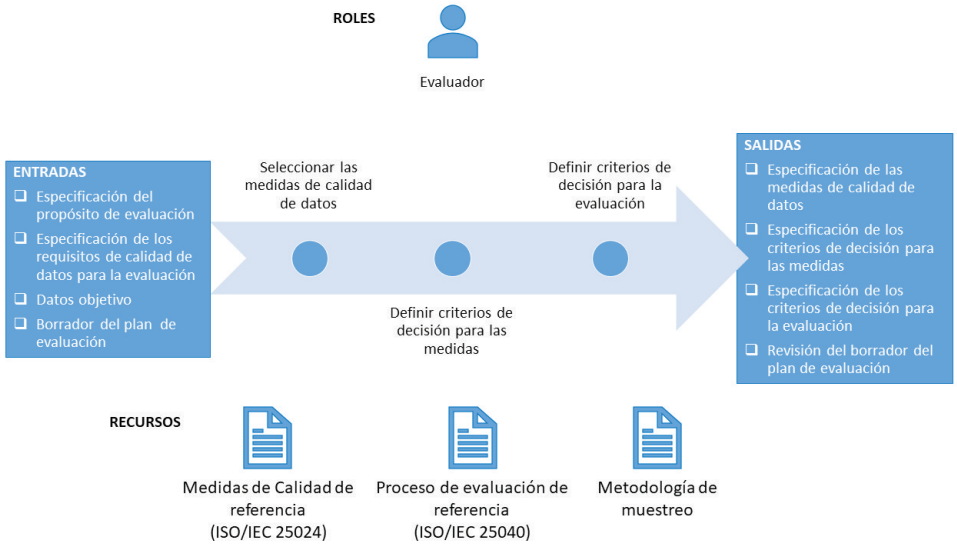

Figura 4.3. Actividad 2 – Especificar la evaluación

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:28.

4.3.2.1 SELECCIONAR LAS MEDIDAS DE CALIDAD DE DATOS (MÓDULOS DE EVALUACIÓN)

En esta actividad se seleccionan las medidas de calidad encargadas de cuantificar el grado en el que se cumplen cada uno de los requisitos de calidad de datos. El evaluador seleccionará medidas de calidad de datos (módulos de evaluación) para cubrir todos los requisitos de calidad de datos. Los requisitos de calidad de datos deben asignarse a los ficheros que forman el producto de datos.

Se requieren medidas rigurosas para hacer comparaciones fiables, ya sea entre diferentes conjuntos de datos o entre distintas versiones de una misma base de datos. Los procedimientos de medición tienen que tener suficiente precisión para permitir que se establezcan criterios de aceptación y se realicen comparaciones.

4.3.2.2 DEFINIR CRITERIOS DE DECISIÓN PARA LAS MEDIDAS DE CALIDAD DE DATOS

Se definirán criterios de decisión para las medidas individuales seleccionadas. Los criterios de decisión se componen de una función de combinación y umbrales numéricos para describir el nivel de calidad del producto de datos en términos de propiedades de calidad de datos.

4.3.2.3 ESTABLECER LOS CRITERIOS DE DECISIÓN PARA LA EVALUACIÓN

Cada uno criterios de decisión para la evaluación puede ser definido en términos de propiedades de calidad de datos individuales o una combinación ponderada de propiedades de calidad de datos. Los resultados de la combinación final deben utilizarse como base para la evaluación y, en su caso, para la certificación de la calidad del producto de datos.

4.3.3 Diseñar la evaluación

Esta actividad, como se puede observar en la Figura 4.4, consta de una sola tarea:

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:28.

- Planificar las actividades de evaluación.

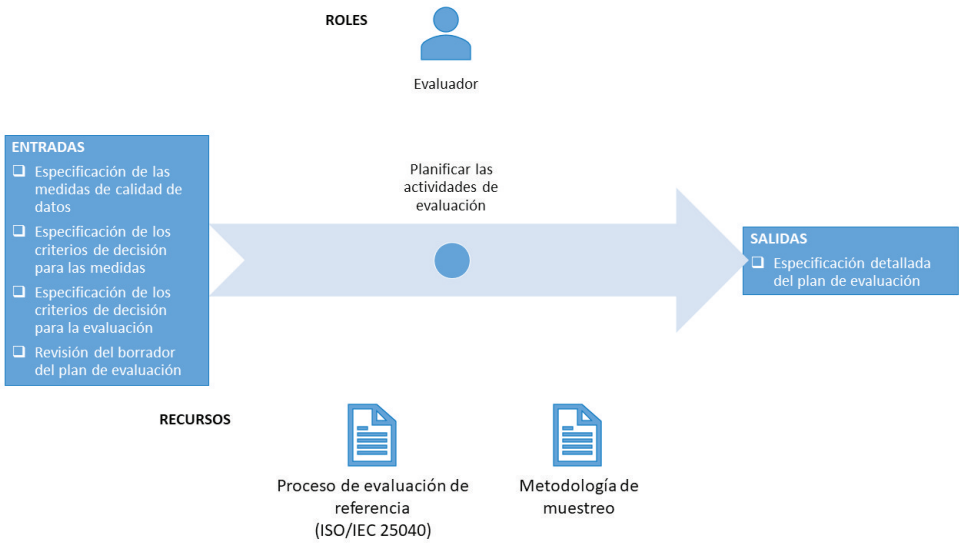

Figura 4.4. Actividad 3 – Diseñar la evaluación

4.3.3.1 PLANIFICAR LAS ACTIVIDADES DE EVALUACIÓN

Las actividades de evaluación de la calidad se programarán teniendo en cuenta la disponibilidad de recursos. Como la naturaleza de la evaluación es iterativa, el plan de evaluación probablemente se modificará a medida que se realice la evaluación.

El plan de evaluación define los puntos de decisión en el proceso de evaluación que determinan cuándo y por qué la evaluación debe considerarse completa (esto es, criterios de aceptación o rechazo) y debe detenerse. Esto debe hacerse con el fin de disminuir el riesgo de errores y reducir el esfuerzo de evaluación considerando al menos los siguientes puntos:

- El presupuesto de evaluación.
- Métodos de evaluación y normas adaptadas.
- Herramientas de evaluación.
- Actividades de evaluación, incluyendo el cronograma y los recursos involucrados.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:28.

El plan de evaluación debe incluir:

- Propósito de la evaluación de la calidad de datos.
- Organizaciones involucradas en la evaluación, como una organización de evaluación independiente o propietarios de productos de datos.
- Presupuesto de evaluación o una referencia al contrato de evaluación.
- Productos de datos a evaluar.
- Calendario para los hitos de la evaluación.
- Responsabilidades de las partes involucradas en la evaluación.
- Métodos y herramientas de evaluación.
- Criterios de decisión para las medidas de calidad de los datos.
- Criterios de decisión para la evaluación de la calidad de datos.
- Normas adoptadas.
- Actividades de evaluación.

Durante los primeros pasos de la evaluación, algunos de estos elementos del plan de evaluación sólo pueden definirse a alto nivel. Por lo tanto, el plan de evaluación se revisará a medida que evolucionen las actividades de evaluación, proporcionando información adicional que permita ajustar o detallar el plan. El plan de evaluación de alto nivel se revisa paso a paso el plan de nivel detallado en las actividades y tareas de evaluación sucesivas.

4.3.4 Ejecutar la evaluación

Esta actividad consta de tres tareas:

- Realizar las mediciones.
- Aplicar los criterios de decisión para las medidas de calidad de datos.
- Aplicar los criterios de decisión para la evaluación de la calidad de datos.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:28.

En la Figura 4.5, se muestran las entradas y resultados, los recursos necesarios y los roles involucrados.

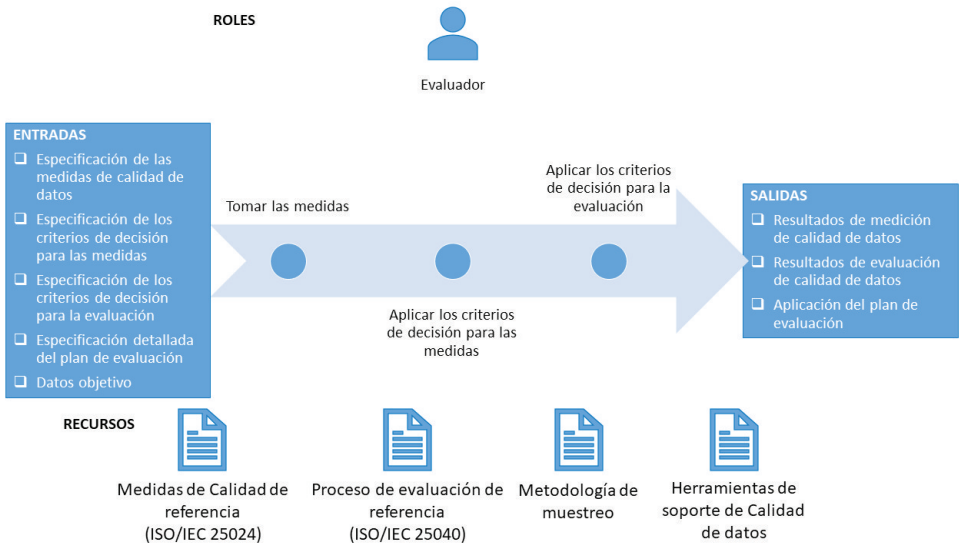

Figura 4.5. Actividad 4 – Ejecutar la evaluación

4.3.4.1 REALIZAR LAS MEDICIONES

Las medidas de calidad de datos seleccionadas se aplicarán a los productos de datos, de acuerdo con el plan de evaluación, obteniendo los resultados de medición.

4.3.4.2 APLICAR CRITERIOS DE DECISIÓN PARA LA EVALUACIÓN

El conjunto de medidas de calidad se agrega en características, produciendo los resultados de la evaluación como una declaración de la medida en que los productos de datos cumplen con los requisitos de calidad de datos. Los resultados de la evaluación deben:

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:28.

1. Establecer un grado de confianza adecuado de que el producto de datos puede cumplir los requisitos de evaluación.

2. Identificar cualquier defecto específico con respecto a los requisitos de evaluación y cualquier evaluación adicional necesaria para determinar el alcance de dichos defectos.

3. Identificar las limitaciones o condiciones especiales que se aplican al uso del producto de datos.

4. Identificar cualquier debilidad u omisión en la propia evaluación y cualquier evaluación adicional que sea necesaria.

5. Identificar cualquier defecto de uso del producto de datos descubierto por la evaluación.

Los resultados de la evaluación se pueden utilizar para apoyar decisiones de gestión, ya que la calidad agregada se puede comparar con otros aspectos como el tiempo y el costo. Las decisiones de gestión incluyen la aceptación o el rechazo, o en el uso o no del producto de datos, entre otros.

4.3.5 Concluir la evaluación

Esta actividad consta de cuatro tareas:

- Revisar los resultados de la Evaluación.
- Crear el informe de evaluación.
- Revisar la calidad de la evaluación.
- Realizar la eliminación de los datos de evaluación.

En la Figura 4.6, se muestran las entradas y los resultados, los recursos necesarios y los roles involucrados.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:28.

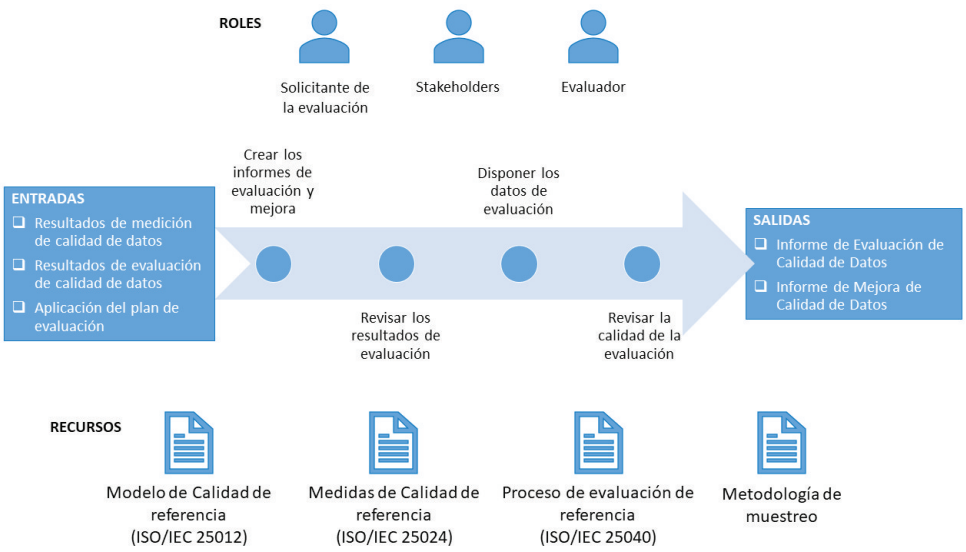

Figura 4.6. Actividad 5 – Concluir la evaluación

4.3.5.1 REVISAR LA EVALUACIÓN DE LA CALIDAD

El evaluador revisará los resultados de la evaluación y la validez del proceso de evaluación, los indicadores y las medidas aplicadas. La retroalimentación de la revisión debe utilizarse para mejorar el proceso de evaluación y las técnicas de evaluación (módulos de evaluación). Cuando sea necesario mejorar los módulos de evaluación, debe incluirse la recogida de datos para indicadores adicionales, con el fin de validarlos para su uso posterior.

4.3.5.2 CREAR EL INFORME DE EVALUACIÓN

Reuniendo todos los resultados y la documentación utilizada, se elaborará el informe de evaluación, que contiene los resultados globales de medición y evaluación.

4.3.5.3 REALIZAR LA ELIMINACIÓN DE LOS DATOS DE LA EVALUACIÓN

Una vez finalizada la evaluación, los datos y elementos de evaluación se eliminarán de acuerdo con los requisitos del solicitante de la evaluación.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:28.

Esto se hará de una de las siguientes maneras, dependiendo del tipo de datos:

- Los documentos presentados a la evaluación serán devueltos al solicitante o archivados durante un tiempo determinado o destruidos de manera segura.
- El informe de evaluación y los registros de evaluación se archivarán durante una duración determinada.
- Todos los demás datos se archivarán durante un tiempo determinado o se destruirán de forma segura.

4.4 CERTIFICACIÓN DE LA CALIDAD DE LOS PRODUCTOS DE DATOS

La certificación es la validación por un tercero (es decir, un organismo de certificación) del cumplimiento de los requisitos previstos por parte del producto de datos.

4.4.1 Entorno de certificación

En la certificación de la calidad de datos, podemos identificar una serie de roles como son:

- **Cliente:** es la parte interesada en la evaluación y certificación de la calidad de los productos de datos propios y solicita tales servicios.
- **Laboratorio de evaluación de la calidad:** es la organización encargada de llevar a cabo la evaluación de la calidad de datos del cliente, y proporcionar el informe de evaluación. Idealmente, deberían ser laboratorios acreditados, en el caso de España por ENAC (Entidad Nacional de Acreditación).
- **Entidad de certificación:** aporta la experiencia en certificación y auditoría. Es la organización encargada de proporcionar el certificado en caso de que el producto de datos evaluado alcance el nivel de calidad necesario.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:43.

- **Consultores:** son las organizaciones encargadas de ayudar al cliente a mejorar la calidad de sus datos para que alcancen el nivel necesario para poder ser certificadas.
- **Distribuidores de herramientas:** de medición o de otro tipo que puedan ayudar a evaluar y mejorar la calidad de los datos.

En la Figura 4.7, se muestra un ejemplo de este entorno, en el que el laboratorio es AQCLab, la entidad de certificación AENOR, y la consultora DQTeam.

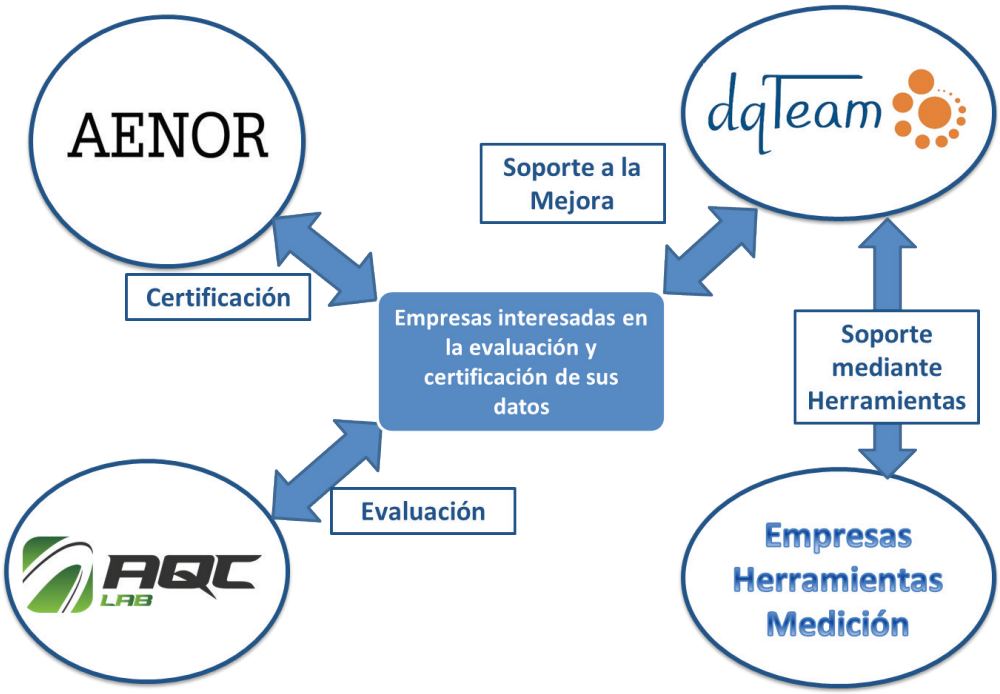

Figura 4.7. Ejemplo de entorno de certificación de calidad de datos

4.4.2 Proceso de certificación

El proceso de certificación consta de los siguientes seis pasos que se pueden observar en la Figura 4.8.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:43.

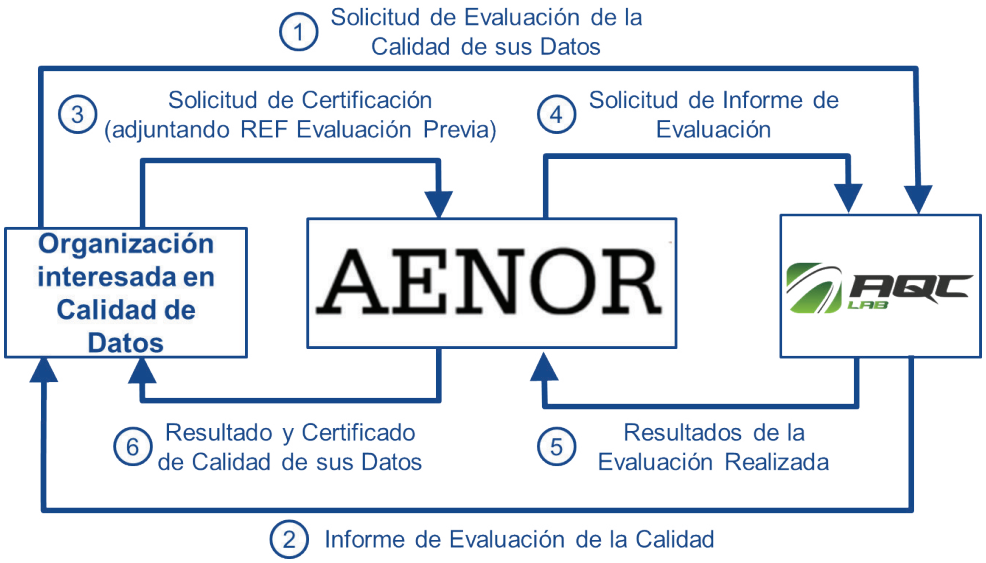

Figura 4.8. Ciclo de certificación

1. El proceso comienza con una organización interesada en la calidad de sus productos de dato. Esta organización solicita una evaluación de la calidad de los datos a un laboratorio de evaluación acreditado –en este caso AQCLab. La organización debe indicar las características del producto de datos a evaluar. El laboratorio AQCLab emite entonces un contrato de evaluación. Cuando el contrato es aceptado y firmado, la organización debe proporcionar todo lo necesario para comenzar la evaluación al laboratorio, incluyendo el producto de datos a evaluar. A partir de este punto, el laboratorio evaluará la calidad del producto de datos.

2. El resultado del paso anterior es un informe de evaluación sobre la calidad del producto de datos. Este informe se proporciona a la organización que solicitó dichos servicios. Para pasar al siguiente paso, AENOR solicita al menos el nivel 3 sobre 5 para la certificación de la calidad de los productos de datos. Si no se obtiene el nivel 3, la organización que solicitó los servicios debe mejorar el producto de datos en términos de su calidad. Una vez que se hayan aplicado esas mejoras, la organización debe solicitar una nueva evaluación de la nueva versión del producto de datos.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:43.

3. Cuando el producto de datos obtenga resultados positivos de evaluación de la calidad de los datos, la organización puede ponerse en contacto con AENOR para solicitar la certificación ISO/IEC 25012 del producto de datos en términos de su calidad. La organización debe hacer referencia a la anterior evaluación favorable. Esta referencia debe incluir las características del producto de datos, la versión del producto de datos y la fecha de la evaluación.

4. AENOR se pondrá en contacto con AQCLab para solicitar los resultados de la evaluación de la calidad del producto de datos referenciado por la organización solicitante en el paso anterior. AENOR confirmará la validez y veracidad de los resultados de la evaluación.

5. AQCLab, como laboratorio de evaluación de calidad, revisará los registros e informes de evaluación para facilitar la información solicitada por AENOR en relación con el producto de referencia.

6. En el último paso, AENOR analiza el informe de evaluación facilitado por AQCLab y audita el producto de datos de la organización. AENOR sigue los reglamentos internos de auditoría de productos de datos, inspección de los productos de datos y sus características. Como resultado de este proceso de auditoría, AENOR emite un informe de certificación y, en caso de resultados positivos, el certificado de calidad del producto de datos (véase Figura 4.9). Tanto el informe de certificación como el certificado incluyen la siguiente información:
   - Información del producto de datos certificado (por ejemplo, versión, fecha, etc.).
   - Características de la calidad de los datos evaluados.
   - Una referencia al informe de evaluación de la calidad de los datos emitido por el laboratorio acreditado de evaluación de la calidad.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:43.

AENOR

Certificado de Conformidad
Calidad de Producto Datos

Producto Datos
ISO 25012
AENOR conform

PD-2016/0001

AENOR certifica que el Producto Datos denominado:

[redacted]

de la empresa

[redacted]

con domicilio social: [redacted]

conforme con: La norma ISO/IEC 25012. Ingeniería de Software. Calidad de Producto Software. Requisitos y Evaluación (SQuaRE) – Modelo Calidad de Datos.

para las características: Precisión, completitud, consistencia, credibilidad, actualidad.

sistema de certificación: Este certificado solo es válido considerando el informe de Evaluación ID: [redacted] con fecha [redacted] del Laboratorio AQCLab acreditado por ENAC y el informe de auditoría de certificación de producto datos de AENOR PD-01/16 con fecha 2016-12-09.

Fecha de primera emisión: 2016-12-09
Fecha de expiración: 2019-12-09

AENOR INTERNACIONAL, S.A.U.
Génova, 6. 28004 Madrid. España
Tel. 91 432 60 00.- www.aenor.com

Avelino BRITO MARQUINA
Director General

Figura 4.9. Ejemplo de certificado de calidad de datos de AENOR.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:43.

4.5 EJEMPLOS DE CERTIFICACIÓN DE CALIDAD DE DATOS

En este apartado presentamos algunos ejemplos basados en casos reales, que nos permitan entender cómo utilizar los elementos explicados en este capítulo para la medición, evaluación y certificación de la Calidad de Datos: el Modelo de Calidad de datos, el Proceso de Evaluación y el Entorno de Certificación.

4.5.1 Modelo de Calidad de Datos

El modelo de calidad de datos contiene los criterios que se deben utilizar para analizar la calidad de los datos. En este ejemplo, el objetivo era la medición de los niveles de calidad de un repositorio de datos con las características indicadas en la Tabla 4.2.

Datos Objetivo	**AlarNet:** Datos sobre miembros y proyectos de un grupo de investigación de la Universidad de Castilla-La Mancha
Tipo de repositorio de datos	Base de datos relacional
Tecnología	Microsoft SQL Server 2003
Ficheros/Tablas	30
Número medio de atributos/columnas por Fichero	15
Número medio de registros/filas por Fichero	60
Reglas de Negocio	92

Tabla 4.2. Características del repositorio de datos del estudio de caso 1

Esta información debe ser proporcionada por el dueño de los datos objetivo y es imprescindible para el análisis:

- El tipo de repositorio de datos y la tecnología para implementar los métodos de medición en el SQL específico de la tecnología.
- El número de ficheros, y los números medios de atributos y registros por fichero para estimar el esfuerzo y planificar un calendario para la medición.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:54.

- Las reglas de negocio para estimar el esfuerzo, y para construir los métodos de medición.

Una vez recibidos las reglas de negocio, los datos (o al menos, el modelo de datos), y conocida la tecnología en la que se han implementado los datos, se puede comenzar con la construcción de los métodos de medición de calidad de datos:

1. En primer lugar, las reglas de negocio deben analizarse para inferir los requisitos de datos (restricciones sobre los datos) que se derivan de ellas. De una regla de negocio pueden inferirse distintos requisitos de datos.

   Ejemplo:

   - Regla de negocio: *"Todos los asalariados de la empresa deben estar dados de alta en la seguridad social"*.
   - De esta regla se extrae que el requisito de datos*: "todo trabajador de la empresa debe tener un número de la seguridad social"*.

2. Los requisitos de datos inferidos de las reglas de negocio se deben clasificar según la característica afectada. Además, hay que identificar qué Propiedad de Calidad de Datos, asociada a las características, se debe utilizar para medir el nivel de calidad de datos.

   Ejemplo: La característica afectada es consistencia y la propiedad que se puede utilizar para medir el nivel de cumplimiento del requisito de datos del ejemplo es Consistencia Semántica.

3. Una vez identificada la Propiedad, se debe desarrollar el método de medición que se debe aplicar para cuantificar los requisitos de datos, y se debe implementar en el SQL específico de la tecnología utilizada para implementar los datos.

En este caso, una vez se realizó esta tarea con todas las reglas de negocio, se planificó la ejecución de los scripts SQL con los requisitos de datos clasificados por Característica y Propiedad de Calidad de datos asociados. Un ejemplo de Script para el requisito de datos anterior se proporciona a continuación:

```
SELECT COUNT(*) FROM Persona WHERE (
tipo_persona IN ('Trabajador', 'Autónomo') AND numero_SS IS NOT NULL
)
```

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:54.

Los scripts SQL se ejecutaron, según la planificación, para la medición de la calidad de los datos. Cada resultado para cada tabla se anotó para su posterior procesamiento.

Los resultados se contrastaron con las reglas de negocio y requisitos de datos asociados. Si una tabla no tiene reglas de calidad asociadas a una propiedad de calidad de datos, no hay valor de medición relacionado.

Los resultados obtenidos para cada tabla según la propiedad se agregaron para unificar el valor de cada Propiedad de Calidad de datos. Los resultados de medición para cada propiedad se combinaron para proporcionar un valor para las distintas Características de Calidad de Datos. Estos valores sirvieron para determinar el estado de la calidad del repositorio de datos con respecto a las reglas de negocio proporcionadas. Un ejemplo de posibles resultados de Consistencia se muestra en la Figura 4.10.

Figura 4.10. Ejemplo de Valores las propiedades de Consistencia de Datos

4.5.2 Proceso de Evaluación de Calidad de Datos

El proceso de evaluación contiene las actividades y las tareas a llevar a cabo para la evaluación de la calidad de los datos. En este ejemplo, el objetivo

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:54.

era la evaluación de la calidad de un repositorio de datos con las características indicadas en la Tabla 4.3, y la identificación de las mejoras posibles.

Datos Objetivo	Datos de una entidad pública
Tipo de repositorio de datos	Base de datos relacional
Tecnología	Oracle Database 11c
Ficheros/Tablas	75
Número medio de atributos/columnas por Fichero	14
Número medio de registros/filas por Fichero	114666
Reglas de Negocio	850

Tabla 4.3. Características del repositorio de datos de estudio de caso 2.

En este caso, tan sólo se proporcionaron en un principio el producto de datos a evaluar y las necesidades de evaluación de calidad de datos, ya que no existía ninguna una versión previa de los requisitos de datos, ni tampoco, una especificación de reglas de negocio definida. En este caso, una consultora se hizo cargo del descubrimiento, elicitación y recolección de las reglas de negocio relacionadas con los datos.

En la primera actividad del proceso de evaluación se estableció su propósito: *"controlar la calidad del producto de datos y detectar defectos de datos para decidir cuándo mejorar su calidad"*. Los requisitos de datos se infirieron de forma análoga al procedimiento llevado a cabo en el estudio de caso 1 (ver apartado 5.1). Durante esta actividad, se identificaron los *stakeholders* relacionados con el repositorio de datos a evaluar: el responsable de los datos, los responsables de cada área de negocio que trabaja con el repositorio de datos, el dueño de los datos y el responsable del tratamiento de los datos.

Durante la actividad de especificación de la evaluación, se utilizó un Modelo de Calidad de Datos para la definición de las medidas y los criterios decisión para la medición de calidad de datos (análogo al procedimiento llevado a cabo en el apartado 5.1). Los criterios de evaluación fueron definidos junto con el responsable de los datos y conforme a las necesidades expresadas por el dueño de los datos.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:54.

En las actividades tercera y cuarta, se planificó y se ejecutó la evaluación respectivamente, teniendo en cuenta las necesidades de evaluación expresadas en la primera actividad del proceso de evaluación.

Por último, la evaluación se concluyó con una revisión de los resultados conjunta con el responsable y el dueño de los datos. Los resultados fueron negativos desde el punto de vista de la evaluación (ver Figura 4.11).

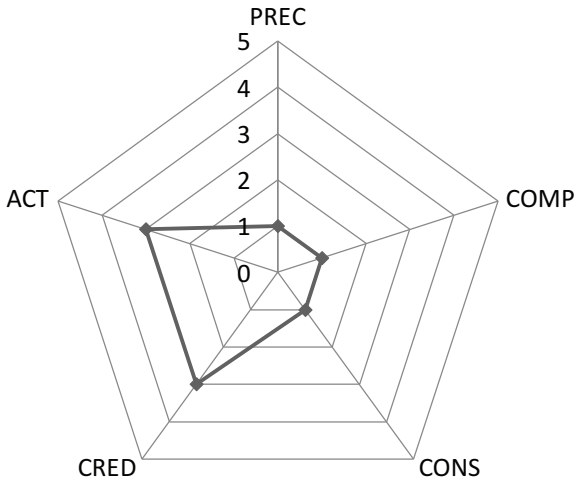

Figura 4.11. Resultados de Medición de Calidad de los Datos de ***PU***

Como resultado no sólo se proporcionó el informe de evaluación, sino además un informe de mejora con la detección e identificación de defectos de datos y una priorización de la resolución de esos defectos. Además, también se identificaron las fuentes más probables de esos defectos, lo que sirvió para resolver los problemas desde el origen para que no volviesen a reproducirse.

4.5.3 Certificación de Calidad de Datos

El entorno de certificación de calidad de datos incluye un proceso para la certificación de datos con respecto a su nivel de calidad y la identificación de roles relacionados con esta tarea de certificación. En este ejemplo, el objetivo era la certificación con respecto al estándar ISO/IEC 25012 de un repositorio de datos con las características indicadas en la Tabla 4.5.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:54.

Datos Objetivo	Datos de usuarios de IED
Tipo de repositorio de datos	Base de datos relacional
Tecnología	Azure SQL Database
Ficheros/Tablas	20
Número medio de atributos/columnas por Fichero	24
Número medio de registros/filas por Fichero	39600
Reglas de Negocio	350

Tabla 4.5. Características del repositorio de datos de estudio de caso 3.

Una vez fueron definidas las responsabilidades de cada uno de los integrantes en la certificación de la calidad del producto de datos del cliente, se inició el proceso de certificación, primero mediante una petición de *IED* a AQCLAb para la evaluación del producto de datos objetivo. AQCLab proporcionó un informe de evaluación siguiendo el proceso de evaluación (ver apartado 3), ejemplificado en el caso de estudio 2 (ver apartado 5.2).

Tras una primera evaluación, se detectó que los niveles de calidad de datos (ver Figura 4.12) eran insuficientes para la certificación del producto de datos objetivo (ver Tabla 5).

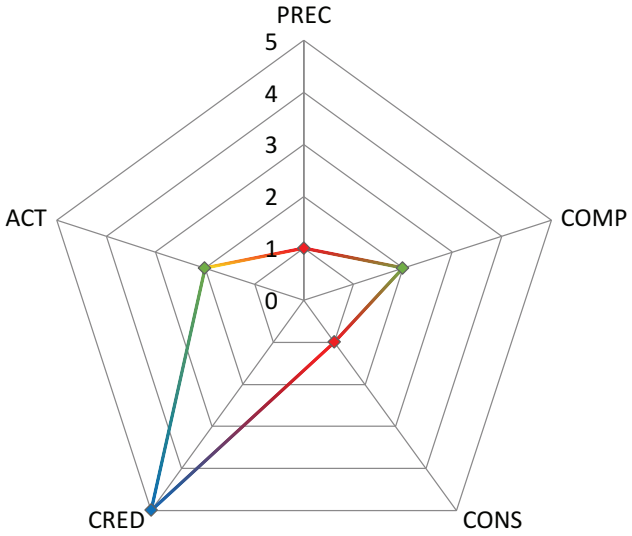

Figura 4.12. Resultados de la primera Evaluación de Calidad de los Datos de *EID*.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:54.

De esta manera, *IED* realizó mejoras en el producto de datos basadas en el informe de evaluación y pudieron detectar los defectos en los datos que afectaban a la calidad del producto de datos objetivo. Tras la limpieza de esos defectos, volvieron a realizar la petición de evaluación a AQCLab. El nuevo informe de evaluación demostraba que los resultados eran positivos para proseguir con la certificación (ver Figura 4.13).

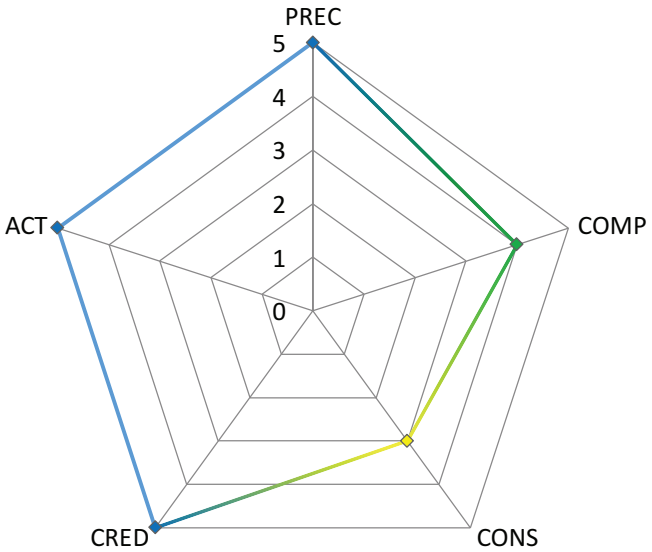

Figura 4.13. Resultados de la segunda Evaluación de Calidad de los Datos de *EID*

Tras obtener el informe con los resultados positivos, *IED* contactó con AENOR para la petición de certificación de la calidad de su producto de datos. AENOR, contactó con AQCLab para conseguir el informe de evaluación y confirmar los resultados de evaluación en una auditoría realizada a *IED*. Los resultados positivos de la auditoría trajeron como consecuencia la emisión del certificado de calidad del producto de datos de *IED*.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:32:54.

4.6 LECTURAS RECOMENDADAS

- Caballero, I., Rodríguez, M., y Fernández, C.M. (2017). Calidad de datos digitales certificada. Revista AENOR. *http://www.aenor.com/revista/pdf/nov17/36nov17.pdf*

  En este artículo se explica la calidad de los productos de datos y se resume la primera experiencia de certificación de calidad de datos basada en ISO/IEC 25012 a nivel internacional.

4.7 SITIOS WEB RECOMENDADOS

- *www.aenor.es*

  Se trata del sitio web de AENOR, entidad que trabaja en los ámbitos de la evaluación de la conformidad y actividades asociadas, como la formación o la venta de publicaciones. AENOR tiene en España presencia en todas las Comunidades Autónomas, a través de 20 sedes, así como presencia permanente en 12 países principalmente de Iberoamérica y Europa.

- *www.aqclab.es*

  AQCLab es el primer laboratorio acreditado para la Evaluación de la Calidad del Producto Software y la Calidad de los Datos en base al estándar internacional ISO/IEC 25000 a nivel internacional.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-28 07:12:33.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-28 07:12:33.

5

MONETIZACIÓN DE LOS DATOS

5.1 INTRODUCCIÓN

Como ya se ha señalado, en los últimos años se repite la idea de que los datos son el nuevo recurso natural mundial en este siglo, al igual que lo fue la máquina de vapor en el siglo XVIII, el electromagnetismo en el XIX, y los combustibles fósiles en el XX[5]. Sorprende, sin embargo, que, siendo los datos tan esenciales para la sociedad y para la transformación digital, no exista un método generalmente aceptado para determinar el valor de los datos y la información para la organización.

Si "los datos son el nuevo petróleo" (Rotella, 2012), merece la pena explorar cómo se valoran los activos relacionados con el petróleo. En este sector, existen valores aceptados, como que el *Brent* se cotiza a 50 dólares/barril, o el *West Texas Intermediate* a 42, o la OPEC *Reference Basket* a 46, … dependiendo de las diferentes "calidades" o tipos de petróleo (más de 161 tipos considerando su densidad, contenido de azufre, etc.). Además, en el mundo real, está claro que no es lo mismo el valor del "crudo" que los combustibles refinados. También en este campo existen diferentes métodos para valorar las reservas de crudo, que además también viene determinadas por su situación geopolítica y los riesgos a las que están sometidas.

Copyright © 2018. RA-MA Editorial. All rights reserved.

5 *http://www.bdva.eu/*

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:46:32.

Sin embargo, en nuestro campo, no existen métodos aceptados que diferencien a la hora de valorar datos de información más o menos elaborada, ni tampoco para valorar bases de datos, repositorios, almacenes de datos o *data lakes*.

El problema es que los especialistas en datos han descuidado el tema del valor del dato, mientras que los economistas y gestores encuentran muchos problemas a la hora de aplicar sus ideas a los datos. Por ejemplo, mientras que los activos "normales" se deprecian con el uso, la mayoría de los datos aumentan su valor con el uso, la información y los datos no son agotables (Otto, 2015). Otro ejemplo ilustrativo es que, en el área de los combustibles, la fracturación hidráulica (*fracking*) ha aumentado considerablemente la extracción de gas y petróleo del subsuelo, disminuyendo su valor; mientras que las tecnologías SMACT han aumentado enormemente la cantidad de datos disponibles, pero aumentando su valor. Así, una vez más tenemos que ser conscientes de que las reglas económicas básicas no son válidas cuando se trata de activos de datos.

5.2 CARACTERÍSTICAS ESPECIALES DE LOS DATOS

Ya Moody y Walsh (1999) propusieron siete "leyes de la información" que gobiernan su comportamiento como un bien económico:

- La información es (infinitamente) compartible.
- El valor de la información aumenta con el uso.
- La información es perecedera.
- El valor de la información aumenta con la exactitud.
- El valor de la información aumenta cuando se combina con otra información (integración).
- "Más" no es necesariamente "mejor".
- La información no es agotable.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:46:32.

Recientemente (Laney, 2018), vuelve a insistir que la información tiene unas características especiales que le permiten generar valor de maneras nuevas e interesantes:

- Es altamente reutilizable.
- Es "líquida" en una amplia variedad de contextos.
- No se considera un activo de capital.
- Cuando se intercambia no se grava (con impuestos).
- Es fácilmente replicable.
- Puede ser transferida de forma fácil e instantánea.
- Tiene beneficios exponenciales.

5.3 APROXIMACIONES A LA VALORACIÓN DE LOS DATOS

En la "clásica" revisión de Repo (1989) se presenta como han estudiado el valor de la información los economistas, investigadores en contabilidad, y científicos de gestión. En esta revisión se destaca que "*la teoría de la información basada en los trabajos de Shannon y Weaver ha fascinado a los economistas así como a otros científicos en diferentes campos de la ciencia… Shannon no estaba interesado en el contenido del mensaje, sino en las probabilidades asignadas por los receptores potenciales del conjunto de mensajes… de manera que tenemos que abandonar la aproximación basada en la teoría de la información y sus derivados y empezar a buscar soluciones en otro sitio*"

En la comunidad de datos y de la información, los investigadores han enmarcado su investigación alrededor de conceptos como "productos de datos" y "activos de datos" (Wang, 1998; Wang et al., 1998). Moody y Walsh (1999) también trataron con el problema de valorar la información, señalando que "*las dificultades de transferir los principios de contabilidad del terreno de lo físico al ámbito de los activos intangibles, como la información. Surgen problemas, por ejemplo, del hecho de que el valor de la información, y lo mismo para los datos, no pueden ser depreciado*".

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:46:32.

En la década de los 2000 se han llevado a cabo varios estudios sobre la contabilidad de los datos como activos intangibles (Atkinson y McGaughey, 2006), pero siempre han tenido en cuenta la "característica especial" de los datos, y es que la mayoría de los datos aumentan su valor con el uso.

Desde 2010 hemos asistido a la aparición de nuevas plataformas (Amazon, Microsoft, etc.) y a la difusión de la computación en la nube (cloud computing), proponiéndose también el Data-as-a-Service (Vu et al., 2012), y la manera de cobrar a los consumidores por la utilización de los activos de datos. Recientemente, Dalessandro et al. (2014) ha propuesto una estimación para un precio justo para datos comprados, basándose en el valor económico de añadir datos incrementables a las aplicaciones de modelado predictivo.

En los últimos cinco años, algunos expertos se han centrado en el valor de datos específicos, como los de medioambiente y sistemas de salud (Laxminarayan y Macauley, 2012). Dyer y Millard (2012) consideran la economía del aprovisionamiento de datos, técnicas de gestión de datos y marcos para el intercambio de datos, integrándolos en el contexto de un marco común para el análisis del valor de los datos ambientales. Barros et al. (2016) proponen una nueva metodología para llevar a cabo en análisis del valor de la información (VOI, *Value Of Information*) dentro de la gestión de un marco para la gestión de embalses.

Más recientemente, se ha investigado la valoración de los datos de carácter personal. Como señalan Spiekermann y Korunovska (2017): "*existe una nueva clase de activos que se comercializa en los mercados de datos, de hecho, muchas empresas consideran sus valoraciones en bolsa en función de los activos de datos que tiene acerca de sus usuarios*"; por lo que proponen que el compromiso y la propiedad "psicológica" se tengan en cuenta a la hora de valorar los datos personales. Saez-Trumper et al. (2014) propuso un modelo que define el valor de un usuario de acuerdo a las acciones que realiza en Facebook. Kassaa et al. (2016) comentan que "*la información personal se está convirtiendo en una mercancía comercializable*", proponiendo un portal de valoración de información personal que se basa en el valor financiero y de mercado. Shen et al. (2016) se centran en los "Big Personal Data" como un tipo nuevo de activo de datos, proponiendo un modelo de tarificación basado en granularidad de tuplas. Sidgman y Crompton (2016) destacan la relación entre el valor de los datos y la preservación de la privacidad, afirmando que "*un enfoque estratégico para la gestión de datos derivado de la valoración de datos debería ayudar a las*

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:46:32.

organizaciones a contener los costes y generar nuevas oportunidades a la vez que se gestionan los riesgos".

Otros enfoques se basan en la satisfacción de los usuarios (Sun et al., 2014), que a su vez se fundamenta en el IS Success Model by De Lone y McLean (1992), También hay estudios que tratan sobre el valor de los sistemas de apoyo a la decisión, como Carton et al. (2016), pero no específicamente con el valor de los datos en sí mismos. Otro trabajo en esta área es el de Schmarzo y Sidaoui (2017), que identifica y prioriza los casos de negocio basándose en las iniciativas de datos y analíticas.

Otra cuestión con la investigación sobre la valoración de los datos tiene que ver con que la mayoría de las propuestas se basan en las consecuencias de los datos para los procesos de negocio, y que "*el entendimiento común es que la calidad de datos sólo es relevante si los datos se usan (Even y Shankaranarayanan, 2007). Sin embargo, en el caso de los datos maestros, es decir, datos que se necesitan para definir sin ambigüedad y que se usan de forma generalizada en la organización a nivel global… la calidad de los datos debe asegurarse para una variedad de procesos de negocio, pero también para su uso "potencial".* (Otto, 2015*).*

5.4 VALOR DE LOS DATOS

De forma análoga a la Ingeniería del Software basada en valor propuesta por Boehm (2003), y desarrollada por varios otros expertos (Biffl et al., 2006), se podría pensar en el caso de los datos en abordar los siguientes temas:

- Elicitación de (requisitos) reglas de negocio de datos basada en valor.
- Arquitectura de información basada en valor.
- Diseño y desarrollo de bases de datos basada en valor.
- Gobierno de datos basado en valor.
- Gestión de riesgos de datos basada en valor.
- Gestión y evaluación de ciberseguridad del dato basada en valor.
- Construcción y explotación de bases de datos, repositorios, datalakes, etc. basadas en valor.

El problema es determinar qué significa el valor, ya que en general "valor" tiene la connotación de medida cuantitativa, aunque no sea necesario expresarlo

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:46:32.

en términos económicos. Algunos autores se refieren a valores no monetarios como número de vidas salvadas, mejoras en la calidad del medioambiente, reducción del coste de seguros, etc. (Laxminarayan y Macauley, 2012).

Siyam et al. (2015) proporciona una revisión sistemática de la literatura acerca del valor y destacan la "contingencia y subjetividad del valor", y proponen que se tome en consideración que:

- El valor es multidimensional.
- El valor es subjetivo, tiene diferente significado para diferentes *stakeholders.*
- El valor es contingente, ya que se realiza si la mix sea apropiada.
- Los enfoques deberían considerar el valor del producto o sistema que se entrega.
- Los enfoques deberían considerar el proceso de valor de principio a fin.
- Los enfoques deberían considerar como se agrega el valor.
- Los enfoques deberían incorporar varias dimensiones del valor.
- Los enfoques deberían considerar más explícitamente el valor del ciclo de vida.

Para determinar el valor de los datos es necesario considerar varios temas que ya se han comentado:

- Diferencias entre datos, información y conocimiento.
- La naturaleza (estructurada, no estructurada, imagen, etc.) de los datos (Batini y Scannapieco, 2016).
- La clase de los datos (datos de proceso de negocio, datos de producto o servicio, datos de clientes, etc.) (Roger, 2016).
- El tipo de datos (transaccional, histórico, datos maestros, metadatos, datos de referencia).

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:46:32.

- El ciclo de vida de datos/información (desde la creación del dato hasta su eliminación) (por ejemplo, ISACA, 2013, Otto, 2015) y la "cadena de valor de big data" (Curry et al., 2014).
- El acceso (cerrado, abierto o bajo licencia), y el origen (interno o externo) de los datos (Kontio et al., 2015).
- El ecosistema de datos (por ejemplo, el ecosistema de *big data*, Cavanillas et al., 2016).
- La tecnología de gestión de datos (sistema de gestión de ficheros, hojas de cálculo, SGBD relacional, motores de *big data*, facilidades en la nube, etc.).
- Los diferentes sectores y dominios (salud, sector público, financiero, energía, transporte, etc.).
- El nivel de calidad de los datos (pragmático, semántico, social, etc.) (Nelson et al., 2012; ISACA, 2013).
- La calidad de los datos (fiabilidad, credibilidad, precisión, etc. (ISO/IEC 25012) y otros parámetros de calidad (Shen et al, 2016).
- Cuestiones de seguridad de ciberdatos.
- Cuestiones de riesgo y auditoría.
- Cuestiones psicológicas.

Por su parte, Short y Todd (2017) proponen que las organizaciones embeban la valoración de datos en sus estrategias siguiendo tres pasos:

1. Hacer explícitas y compartibles las políticas de valoración de datos en toda la organización.
2. Crear experiencia de valoración de datos dentro de la organización.
3. Decidir si son más efectivos en la organización los procesos de valoración descendentes (en los que se identifican las aplicaciones críticas y se asigna valor a los datos utilizados en las mismas) o ascendentes (de define el valor de los datos a partir de los conjuntos de datos utilizados en la organización).

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:46:32.

5.5 INFONOMÍA

Como se señala en SAS (2017) existen algunas metodologías propietarias que se pueden utilizar para evaluar el valor de la información como la TEI (*Total Economic Impact*) de Forrester, "*Unlocking the Value of Information*" de SAS y Shark Finesse, "Information Management Value Model" de los National Archives, "Customer Data Integration" de Data Connects/Tribal, etc. pero quizás la más desarrollada sea la basada en la Infonomía que resumimos a continuación.

"Infonomía", deriva de "información" y "economía", concede especial atención a los valores de datos y se define como "*la teoría, estudio, disciplina de asignarle importancia económica a la información*" (Laney, 2018), aplicando tanto principios como prácticas económicas y de gestión de activos a la valoración, gestión y despliegue de los activos de información. También se define Infonomía como "la gestión inteligente de la información"[6], ya que pretende dar a individuos y organizaciones ideas y métodos para el mejor uso de la información.

Por tanto, como señala Laney (2018), la infonomía proporciona un marco a las organizaciones para monetizar, gestionar, y medir la información como un activo real. Este autor señala como la monetización puede ser tanto indirecta (que incluye: comerciar con la información, productos o servicios mejorados con información; vender datos sin elaborar a *brókeres* de datos o terceras partes, desarrollar y ofrecer subscripciones a datos/informes (predefinidos o personalizados), o vender soluciones analíticas usando datos como un componente de una solución global), como indirecta (que comprende: mejora de eficiencia, reducciones de riesgo, desarrollo de nuevos productos y mercados, o construcción y consolidación de relaciones con socios. También señala 12 drivers para la monetización:

1. Incrementar la adquisición/retención de clientes.
2. Crear una fuente de ingresos suplementaria.

Copyright © 2018. RA-MA Editorial. All rights reserved.

6 http://www.documentalistaenredado.net/545/que-es-la-infonomia/

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:46:32.

3. Introducir una nueva línea de negocio.
4. Entrar en nuevos mercados.
5. Permitir una diferenciación competitiva.
6. Intercambiar por bienes y servicios.
7. Conseguir términos y condiciones favorables y relaciones mejoradas.
8. Sufragar el coste de la gestión y analítica de la información.
9. Reducir los costes de mantenimiento, los excesos de coste y los retrasos.
10. Mejorar el bienestar de los ciudadanos.

El marco propuesto por Laney (2018) es uno de los más actuales y completos, y está compuesto de tres partes:

- Monetizar (justificación, inspiración y ejecución).
- Gestionar (Barreras, marco, y organización/roles).
- Medir (calidad, valor, economía).

A la hora de monetizar la información, propone los siguientes pasos, que debe ser adaptados a las necesidades y cultura específicas:

- Establecer una función de gestión del producto información encargada de generar beneficios económicos medibles a partir de los activos de información disponibles. Esta función puede reportar al CDO o al responsable analítico.

- Desarrollar y mantener un inventario de los activos de información posibles en la organización, así como de fuentes de segundas y terceras partes, incluyendo datos operacionales, datos "oscuros" (*dark data*) -que son aquellos que no están accesibles o no se encuentran lo suficientemente estructurados para su análisis-, datos comerciales, datos públicos, datos de medios sociales, contenido de la web, etc.

- Evaluar alternativas tanto para la monetización de la información directa como indirecta.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:46:32.

- Identificar, adoptar y adaptar ideas de monetización de la información a alto nivel de otras organizaciones, especialmente aquellas en otras industrias.
- Probar la viabilidad de las ideas de monetización de la información, La viabilidad puede ser: práctica, comercializable, escalable, gestionable, tecnológica, económica, legal, ética y ecológica.
- Preparar y empaquetar la información para la monetización, lo que puede implicar: agregarla, integrarla, complementarla, analizarla, limpiarla, anonimizarla, etc.
- Establecer y cultivar un mercado para los activos de información.

Como modelos de valoración de los activos de datos, Laney (2018) destaca dos tipos de enfoques:

- **Enfoque fundamental**, que considera el aspecto de la información relacionado con la calidad o su impacto en indicadores de desempeño. Dentro de este enfoque consideran tres medidas:
  - Valor intrínseco de la información (IVI, *Intrinsec Value of Information*), que es el supuesto beneficio que permite comparar diferentes clases de información, sin tener en cuenta cómo se utiliza. La fórmula propuesta es:

$$IVI = Validez * Completitud * (1 - EscasezDatos) * CicloDeVida$$

    Donde *CicloDeVida* representa la longitud de uso razonable del activo de información y *EscasezDatos* es la tasa del mercado o de los competidores que probablemente también tengan esos datos.
  - Valor de negocio de la información (BVI, *Business Value of Information*), que es la utilidad de la información para el negocio.

$$BVI = (\sum_{p=1}^{n} relevancia_p) * validez * completitud * oportunidad$$

    Donde p es el número de procesos de negocio o funciones organizacionales.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:46:32.

- Valor de desempeño de la información (PVI, *Performance Value of Information*), que tiene en cuenta el impacto del activo de información en el objetivo de negocio que se representa por medio de KPI (*key performance indicators*).

$$PVI = \left[\left(\frac{KPI_i}{KPI_C}\right) - 1\right] * T/t$$

Donde

- KPI_i son las instancias de procesos de negocio que usan el activo de información.
- KPI_c = son las instancias de procesos de negocio que no usan el activo de información.
- T = la duración media útil de la vida de una instancia de datos.
- t = la duración en la que se ha medido el KPI.

◤ **Enfoque financiero**, miden el valor en términos monetarios adaptando los métodos aceptados para valorar los activos tradicionales. Dentro de este enfoque consideran tres medidas:

- Coste de valor de la información (CVI, *Cost Value of Information*), valora el activo de información como el gasto económico requerido para generar, capturar y recolectarlo.

$$CVI = \frac{ProcExp * Attrib * T}{t}\{+\sum_{p=0}^{n} LostRevenue_p\}$$

En el que:

- ProcExp = es el coste anualizado de los procesos involucrados en capturar el dato.
- Attrib = el porcentaje de gastos del proceso atribuible a la captura del dato.
- T = duración media de la vida de cualquier instancia de datos.
- t = período de tiempo sobre el que se mide el gasto del proceso.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:46:32.

  - n = número de períodos de tiempo hasta que la información se vuelve a adquirir, o hasta que la continuidad del negocio no se ve más afectada por la pérdida o daño de la información.

- Valor de mercado de la información (MVI, *Market Value of Information*), que tiene en cuenta el valor potencial o real de un activo de información en el mercado abierto. Se considera que la información no se vende como tal, sino que se vende como licencia., por lo que se representa como un factor de descuento (la inversa de un recargo) aplicado a una transferencia hipotética de la propiedad (precio exclusivo) del activo de información.

$$MVI = \frac{PrecioExclusivo * NúmeroLicencias}{Recargo}$$

- Valor económico de la información (EVI, *Economic Value of Information*), que genera un valor económico neto de un activo de información aplicando el enfoque tradicional de ingresos para cualquier activo y restando los gastos del ciclo de vida de la información.

$$EVI = [Ingresos_i - Ingresos_c - (GastosAdq + GastosAdm + GastApli)] * T/t$$

Donde

  - Ingresos i = los ingresos generados usando el activo de información
  - Ingresos c = los ingresos generados no usando el activo de información
  - T = la duración media de la vida esperada de cualquier instancia de información
  - t = el período de tiempo durante el cual se ejecuta el cálculo del EVI

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:46:32.

5.6 LECTURAS RECOMENDADAS

- Ontiveros, E. y López, V. (eds.) (2017). Economía de los Datos. Riqueza 4.0. Madrid, Fundación Telefónica.

  Este monográfico trata diversos aspectos relacionados con la economía de los datos: mercados de los datos, oportunidades (incluyendo nuevas profesiones), y principales retos.

- Laney, D.B. (2018). Infonomics. How to monetize, manage, and measure information as an asset for competitive advantage. Nueva York, EE.UU. Bibliomotion.

  Este libro presenta los principales temas relacionados con la infonomía desde un punto de vista práctico.

5.7 SITIOS WEB RECOMENDADOS

- *www.gartner.com*

  En el portal de Gartner se presentan periódicamente diversos informes y experiencias sobre la monetización de los datos.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:46:32.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-27 09:46:32.

Anexo I

PROBLEMAS DE CALIDAD DE DATOS EN BASES DE DATOS RELACIONALES

Según Oliveira et al. (2005), los problemas de calidad de datos en bases de datos relacionales se pueden tratar cuatro niveles de granularidad: varias fuentes de datos, varias relaciones (tablas), una simple relación, y nivel de atributo/tupla. En esta clasificación de problemas, los autores no incluyen datos multimedia.

A1.1 PROBLEMAS A NIVEL DE ATRIBUTO / TUPLAS

Este tipo de problemas se pueden dar en los siguientes elementos: i) un atributo de una tupla; ii) un atributo en muchas tuplas (a nivel de columna); o iii) varios atributos a nivel de una tupla (a nivel de fila).

A1.1.1 Un atributo en una tupla

- Valores perdidos, por ejemplo, cuando falta un valor en un atributo en cuya definición se estableció como no nulo. Este tipo de problema tiene que ver con la dimensión de la completitud.
- Violación de sintaxis, cuando por ejemplo se establece una expresión regular para el formato de un atributo de correo electrónico de la forma *usuario@subdominio.dominio* y se almacena "*ismael@uclm*". Este tipo de problema tiene que ver con la dimensión de la precisión sintáctica.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:51:15.

- Valores incorrectos, por ejemplo, cuando en una base de datos se tiene el atributo nombre autor, y se sabe que es "*Mario Piattini*", aparece "*Mariano Piattini*". Este tipo de problema tiene que ver con la dimensión de la precisión semántica.

- Violación de dominio, por ejemplo, cuando para el atributo *NúmeroOrden* aparece el valor -5 (en una secuencia de orden, no deben aparecer valores negativos).

- Violación de las restricciones, por ejemplo, cuando se establecen que para poder registrarse en una web hace falta ser mayor de edad y no se verifica que (*EdadActual >=18*).

- Subcadena con contenido inválido para el tipo de dato, se correspondería con un tipo de violación de dominio, por ejemplo, cuando al almacenar el código postal, se almacenará también el nombre de la localidad.

- *Error de Deletreo,* es un tipo de error específico que se correspondería con una violación de sintaxis cuando hay una variación de una letra en el valor del atributo. Por ejemplo, cuando se almacena "*Ismeal*" por "*Ismael"*.

- Valores imprecisos, es un tipo de error específico cuando no es posible interpretar un valor sustitutivo que se da. Por ejemplo, cuando se especifica "ISO", y no se sabe si se corresponde con "*International Organization for Standardization*" o con "*Inspection of Supplies*".

A1.1.2 Atributos en varias tuplas (alcance de columna)

- Violación de valor único. Ocurre cuando por ejemplo dos personas diferentes se almacenan teniendo el mismo DNI.

- Existencia de sinónimos. Este problema aparece cuando aparecen como valores del atributo palabras que son consideradas como sinónimas. Por ejemplo, cuando para el atributo Profesión en una tupla aparece "Profesor" y en otras aparece "Maestro".

- Violación de restricciones, que afecten a varias tuplas.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:51:15.

A1.1.3 Varios atributos en una tupla

- Tupla semivacía. Ocurre cuando no todos los atributos de una tupla tienen valores. Para considerar que una tupla puede seguir siendo válida se puede identificar un valor umbral del número de atributos que no deben ser nulos (e.g. 60%), o incluso, en función de la importancia de los atributos, dar un peso específico a cada uno de los atributos a la hora de computar la validez de la tupla con respecto al valor umbral.

- Violación de dependencia funcional. Supóngase una tabla que contenga los atributos, *CódigoPostal* y *Ciudad*, de los que se sabe que *CódigoPostal* determina *Ciudad*. Se produce una violación de dependencia funcional cuando en la tabla aparecen para una tupla *CódigoPostal=13071* y *Ciudad= "Ciudad Real"* y *CódigoPostal= 13071* y *Ciudad= "Madrid"*.

- Violación de restricciones de dominio. Este tipo de violación puede ocurrir cuando se define el valor de un atributo en función de otro(s) dentro de la misma tupla según las restricciones del dominio, pero luego el valor que se inserta no está calculado según esas reglas. Por ejemplo, cuando en una relación (*NumeroUnidades, PrecioUnitario, TotalPrecio)* existe una cantidad que se llama *TotalPrecio* que se calcula como *NúmeroUnidades* * *PrecioUnitario* y a la hora de insertar una tupla se inserta como (4, 30, 150).

A1.2 PROBLEMAS A NIVEL DE RELACIÓN

- **Tuplas aproximadamente duplicadas**. Este problema surge cuando dos o más tuplas tienen valores correspondientes a la misma entidad de la realidad, aunque estos valores no son exactamente iguales. Por ejemplo, cuando aparecen para una relación *Cliente* (*CodPersona, Nombre, Ciudad)* dos tuplas (1223, "Francisco Sánchez", "Ciudad Real") y (8553, "Fco. Sánchez", "Cdad. Real"). Para resolver este tipo de problema se usan las técnicas de resolución de entidad (*Talburt 2011*).

- **Tuplas inconsistentemente duplicadas**. Este tipo de problema es similar al anterior, pero con la diferencia de que, en este caso, las claves primarias son exactamente iguales. Siguiendo el ejemplo anterior, y

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:51:15.

suponiendo que el primer atributo es la clave primara, un ejemplo de Tuplas inconsistentemente duplicadas sería (1223, "Francisco Sánchez", "Ciudad Real") y (1223, "Fco. Sánchez", "Ciudad Real").

- **Violación de restricciones de dominio**. Un ejemplo de situación de este tipo de problema es cuando por una cierta regla de negocio, sólo se permite a un cliente adquirir 10 productos, y en la tabla de *ClienteCompraProducto* aparece que ha comprado ese producto más de 10 veces.

Muchos de los problemas anteriores se pueden afrontar haciendo un diseño más comprensivo o completo del esquema de la base de datos, y además debe ocurrir que el Sistema Gestor de Bases de Datos (SGBD) lo soporte.

A1.3 PROBLEMAS A NIVEL DE VARIAS RELACIONES

- **Violaciones de integridad referencial**. Este es uno de los problemas más típicos en bases de datos relacionales en las que no se definen bien las claves ajenas. Un ejemplo de manifestación de este tipo de problema es cuando se tiene en una tabla una clave ajena cuyo valor no está en la clave primaria a la que referencia. Por ejemplo, cuando se tienen las relaciones *Domicilio (CodPersona, Calle, CP)* y *Poblacion (CP, NombrePoblación)* con *Domicilio.CP* → *Población*, y se tiene una tupla Domicilio ("Ismael", "c\ Paz", 13005), pero en la tabla Población no aparece ninguna tupla con CP=13005.

- **Referencias incorrectas**. En el ejemplo anterior, este tipo de problema aparece cuando, existiendo un posible valor para el CP, éste no se corresponde con el que se pretende almacenar. Por ejemplo, cuando la tupla anterior *Domicilio ("Ismael", "c\ Paz", 18025)*, el valor en realidad para el código postal debería ser 13005.

- **Heterogeneidad de sintaxis**. Este tipo de problema se produce por ejemplo cuando el formato del mismo dato en tuplas diferentes tiene diferentes formatos. Por ejemplo, cuando en la relación *Vuelo (CodVuelo, Origen, Destino, HoraSalida, HoraLlegada)* la hora de salida se expresa como "22:00 CET", mientras que en la relación *Facturacion (CodCliente, CodVuelo, HoraSalida)*, la hora de salida

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:51:15.

se muestra como "10:00 PM CET": son la misma hora, pero como se expresan de forma diferente, puede originar errores en una posible computación.

- **Bucles entre tuplas auto-referenciadas**. Este tipo de problema se suele dar en relaciones que dependen de otras que contienen a su vez referencias a la primera. Por ejemplo, supóngase que se tiene una relación *Producto (CodProducto, CodSubproducto),* donde CodSubproducto hace referencia a un producto que forma parte de otro producto. Puede aparecer este tipo de problemas si dentro de la base de datos existieran dos tuplas Producto ("X", "Y") y Producto ("Y", "X"), es decir, que un producto "X" tiene como subproducto a otro "Y" del que al mismo tiempo es el producto cuyo subproducto es "X".

- **Violaciones de restricciones de dominio**. Siguiendo con el ejemplo de las facturas, este tipo de problema se da cuando el atributo *FacturaciónTotal* de una relación *Facturas (CodCliente, CodFactura, FacturaciónTotal)* tiene el valor 1000€, pero la suma de todas las ventas de la relación *Ventas (CodCliente, CodFactura, CodProducto, Unidades, Precio)* que se facturan en una misma factura tiene un valor calculado de 1200€.

A1.4 PROBLEMAS A NIVEL DE MÚLTIPLES FUENTES

Para presentar este tipo de problemas, asumimos que se tienen dos relaciones que pertenecen a dos fuentes de datos diferentes. Ambas relaciones representan la misma entidad del mundo real, y supongamos, que, por simplicidad, se han resuelto algunos asuntos como heterogeneidad de nombres, aunque el número de atributos en cada una de ella puede variar, incluso, a nivel de granularidad para mantener los mismos conceptos. En estas circunstancias (*Oliveira, Rodrigues et al. 2005*) identifican los siguientes problemas:

- **Heterogeneidad de sintaxis**. Este problema ocurre cuando la sintaxis de un atributo de una relación de una fuente de dato es diferente a la del mismo atributo de otra fuente de datos. En el ejemplo de los vuelos, una aerolínea puede almacenar que la hora de salida de un vuelo son las "*10:00 PM CET*", mientras que, para la gestora del aeropuerto, la

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:51:15.

hora de salida de ese mismo vuelo pueden ser las "22:00 CET". El valor del dato es el mismo, pero está expresado con diferente sintaxis.

- **Heterogeneidad de unidades de medida**. Siguiendo el mismo ejemplo de los vuelos, se produce un problema de este tipo, cuando el precio del vuelo se expresa para una aerolínea en euros, y para una agencia de viaje en dólares.

- **Heterogeneidad de presentación**. Este tipo de problema ocurre cuando en ambas fuentes de datos se eligen representaciones diferentes para el valor de un determinado atributo. Por ejemplo, cuando para el género, en una fuente de datos se elige ("H", "M") y en la otra (0, 1).

- **Existencia de sinónimos**. Se suele dar este tipo de problema cuando una de las fuentes de datos elige nombres diferentes para el mismo concepto, bien por temas culturales, bien por temas de nomenclatura. Por ejemplo, cuando en una relación de personal, el atributo *CategoríaProfesional* para una aerolínea española contiene el valor de *"azafata",* mientras que para una mexicana contiene el valor "*aeromoza"*.

- **Tuplas aproximadamente duplicadas**. Al igual que lo que sucedía para el caso de una única tupla (véase subapartado anterior), puede ocurrir que para una fuente de datos se tenga unos valores, mientras que para la otra fuente se tengan otros similares, pero no exactamente los mismos. Por ejemplo, para el caso de un trabajador en la compañía aérea Iberia se tenga en la relación *Trabajador(CodTrabajador, Nombre, FechaIncorporación)* una tupla ("A001", "Francisco Ruiz", "20/02/1990") y para la asesoría que gestiona las nóminas de los trabajadores, se tiene una relación *Trabajador(CodTrabajador, Nombre, FechaIncorporación)* una tupla ("*Iberia-A001", "Francisco Ruiz", "20/02/1990")*. Obviamente se refieren al mismo trabajador, pero internamente cada organización lo identifica de forma diferente.

- **Tuplas inconsistentemente duplicadas**. Sería una situación especial del caso anterior donde algunos de los atributos no están adecuadamente replicados. Por ejemplo para el caso anterior del trabajador de la compañía aérea Iberia se tiene en la relación *Trabajador (CodTrabajador, Nombre, FechaIncorporación)* una tupla ("A001", "Francisco Ruiz", "20/02/1990") y para la asesoría que gestiona

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:51:15.

las nóminas de los trabajadores, se tiene una relación *Trabajador (CodEmpresa, CodTrabajador, Nombre, FechaIncorporación)* una tupla ("*Iberia", "X230989", "Francisco Ruiz", "20/02/1990").*

- **Violaciones de restricciones de dominio**. Este tipo de problema puede aparecer cuando se han definido unas reglas de negocio que no coinciden en ambas fuentes de datos, pero hacen referencia al mismo hecho. Por ejemplo, cuando en la compañía la suma de todos los equipajes facturados no debe superar los 32 Kg., mientras que para el sistema de facturación del aeropuerto solo permite 23 Kg.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:51:15.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:51:15.

ACRÓNIMOS

- AENOR Asociación Española de Normalización y Certificación
- ALEI Authoritative Legal Entity Identifiers
- ANS Acuerdo de Nivel de Servicio
- AP Atributo de Proceso
- BI Business Intelligence
- BPM Business Process Management
- BVI Business Value of Information
- CDI (Customer Data Integration)
- CDO Chief Data Officer
- CEM Customer Experience Management
- CEO Chief Executive Officer
- CIO Chief Information Officer
- CISA Certified Information Systems Auditor
- CISM Certified Information Security Manager
- CMMI Capability Maturity Model Integration
- COBIT Control OBjectives for Information and related Technology
- CRM Customer Relationship Management
- CVI Cost Value of Information

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:51:29.

- DAMA Data Management Association
- DCAM Data Management Capability Assessment Model
- DG Data Governance
- DM Data Management
- DMBOK, Data Management BOdy of Knowledge
- DMM Data Maturity Management
- DMS Data Management Strategy
- DNI Documento Nacional de Identidad
- DO Data Operation
- DQ Data Quality
- DQM Data Quality Management
- ECCMA Electronic Commerce Code Management Association
- EDM Enterprise Data Management
- EIM Enterprise Information Management
- ERP Enterprise Resource Planning
- ESB Enterprise Service Bus
- EVI Economic Value of Information
- GPDR Reglamento General de Protección de Datos
- ICIQ International Conference on Information Quality
- IEC International Electrotechnical Commission
- ISACA Information Systems Audit and Control Association
- ISO International Standarisation Organization
- IoT Internet of Things
- IT Information Technology
- IVI Intrinsec Value of Information
- JTC Joint Technical Committee
- KPI Key Performance Indicator

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:51:29.

- MAMD Modelo Alarcos de Madurez/Mejora de Datos
- MDM Master Data Management
- MIT Instituto Tecnológico de Massachusetts
- MVI Market Value of Information
- OPEC Organization of the Petroleum Exporting Countries
- OT Operational Technology
- PA Platform & Architecture
- PA Proceso Auxiliar
- PDCA Plan-Do-Check-Act
- PDQ Product Data Quality
- PIM Product Information Master
- PMP Project Management Professional
- POSMAD Plan, Obtain, Store & Share, Maintain, Apply, Dispose
- PP Proceso Principal
- PVI Performance Value of Information
- QA Quality Assurance
- RRHH Recursos Humanos
- SC Subcommittee
- SCM Supply Chain Management
- SDLC System Development Life Cycle
- SEI Software Engineering Institute
- SGBD Sistema de Gestión de Base de Datos
- SMAC Social, Mobile, Analytics, Cloud
- SMACT Social, Mobile, Analytics, Cloud, Internet of Things
- SP Support Process
- SQuaRE Software product Quality Requirements and Evaluation
- TC Technical Committee

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:51:29.

- TEI Total Economic Impact
- TI Tecnologías de la Información
- TIC Tecnologías de la Información y las Comunicaciones
- TOGAF The Open Group Architecture Framework
- TQdM Total Quality data Management
- TS Technical Specification
- TSI Tecnologías y Sistemas de Información
- UCLM Universidad de Castilla-La Mancha
- VOI Value Of Information
- WG Working Group

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central,
http://ebookcentral.proquest.com/lib/bibliotecaupsp/detail.action?docID=5758915.
Created from bibliotecaupsp on 2020-02-27 09:51:29.

BIBLIOGRAFÍA

Aiken, P. (2016). EXPERIENCE: Succeeding at Data Management – BigCo Attempts to Leverage Data. ACM Journal of Data and Information Quality, Vol. 7, Nº 1-2, art. 8, mayo 2016.

Aiken, P., Allen, M. D., Parker, B., y Mattia, A. (2007). Measuring data management practice maturity: A community's self-assessment. Computer, 40(4).

Aiken, P. y Harbour, T. (2017). Data Strategy and the Enterprise Data Executive. Technics Publications, LLC,, USA.

Allen, M. y Cervo, D. (2015). Multi-Domain Master Data Management. Waltham, MA, EE.UU., Morgan Kaufmann.

Akerlof, G.A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. Quarterly Journal of Economics, 84 (3), 488-500.

Atkinson, K. y Mc Gaughey, R. (2006). Accounting for data: A shortcoming in accounting for intangible assets. Academy of Accounting and Financial Studies Journal, 10(2), 85–95.

Barker, J.M. (2016). Data Governance: The Missing Approach To Improving Data Quality. Doctor of Management in Organizational Leadership with a Specialization in Information Systems and Technology. University of Phoenix

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:33:09.

Barros, E. G. D., Van den Hof, P. M. J. y Jansen, J.D. (2016). Value of information in closed-loop reservoir management. Computing Geoscience 20:737–749 DOI 10.1007/s10596-015-9509-4

Batini, C. y Scannapieco, M. (2016). Data and Information Quality. Dimensions, Principles and Techniques. Switzerland, Springer.

Berson, A. y Dubov, L. (2011). Master Data Management and Data Governance, 2ª ed. Nueva York, EE.UU., Mc Graw-Hill.

Biffl, S., Aurum, A., Boehm, B., Erdogmus, H. y Grünbacher, P. (2006). Value-Based Software Engineering. Springer.

Boehm, B. (2003). Value-Based Software Engineering. ACM SIGSOFT Software Engineering Notes 28 (2), 1-7.

Brown, S.A., Venkatesh, V., Kuruzovich, y Massey, J. A.P. (2008). Expectation confirmation: an examination of three competing models, Organizational Behavior and Human Decision Processes 105, 52–66.

Calero, C., Ruiz, R., Baroni, A., Brito e Abreu, F. y Piattini, M. (2006a). An ontological approach to describe the SQL: 2003 object-relational features. Computer Standards & Interfaces 28(6): 695-713.

Calero, C., Ruiz, F. y Piattini, M. (eds.). (2006b). Ontologies in Software Engineering and Software Technology. Springer.

Cappiello, C., et al. (2010). "Information Quality in Mashups." Internet Computing, IEEE 14: 14-22.

Carretero, A. G., Caballero, I., y Piattini, M. (2016a). Evaluación del nivel de madurez de datos usando MAMD: un estudio de caso. Paper presented at the Congreso Iberoamericano en Ingeniería de Software (CibSE).

Carretero, A. G., Caballero, I., y Piattini, M. (2016b). MAMD: Towards a data improvement model based on ISO 8000-6X and ISO/IEC 33000. International Conference on Software Process Improvement and Capability Determination, 241-253.

Carretero, A. G., Gualo, F., Caballero, I., y Piattini, M. (2017). MAMD 2.0: Environment for data quality processes implantation based on ISO 8000-

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:33:09.

6X and ISO/IEC 33000. Computer Standards and Interfaces, 54, 139-151. doi:10.1016/j.csi.2016.11.008.

Carton, F., Hynes, T. y Adam, F. (2016) A business value oriented approach to decision support systems, Journal of Decision Systems, 25:sup1, 85-95, DOI: 10.1080/12460125.2016.1187415.

Cavanillas, J.M., Curry, E., y Wahlster, W. (2016). New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe. *http://link.springer.com/book/10.1007/978-3-319-21569-3*.

CMMI (2014). Data Management Maturity Model. Versión 1.0. CMMI Institute.

Chandrasekaran, N. (2015). Is data the new currency? World Economic Forum, 14 August 2015. https://www.weforum.org/agenda/2015/08/is-data-the-new-currency/.

Cordon, S., Garcia-Mlià, P., Ferreiro, T. y Caballero, P. (2016). Strategy is digital. How Companies Can Use Big Data in the Value Chain. Switzerland, Springer.

Creswell, J. (2014). Research design: Qualitative, Quantitative and Mixed Methods Approaches. 4th ed. Sage Publications.

Curry, E., Ngonga, A., Domingue, J., Freitas, A., Strohbach, M., Becker, T., et al. (2014). D2.2.2. Final version of the technical white paper. Public deliverable of the EU-Project BIG (318062; ICT-2011.4.4).

Dalessandro, B., Perlich, C. y Raeder, T. (2014). Bigger Is Better, But at What Cost?. Estimating the Economic Value of Incremental Data Assets. Big Data 2 (2). DOI: 10.1089/big.2014.0010.

DAMA. (2017). DAMA-DMBOK: Data Management Body of Knowledge (2ª Ed.). Bradley Beach, Nueva Jersey: Technics Publications.

De la Peña, J. y Cabezas, M. (2015). La Gran Oportunidad. Claves para liderar la transformación digital en las empresas y en la economía. Barcelona, Gestión 2000.

De Lone, W.H. y McLean, E.R. (1992). Information systems success: the quest for the dependent variable, Information Systems Research 3, 60–95.

Dobbs, R., Manyika, J. y Woetzel, J. (2015). No Ordinary Disruption. Public Affairs, Nueva York, EE.UU.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:33:09.

Dyer, B. y Millard, K. (2012). A generic framework for value management of environment data in the context of integrated coastal zone management. In Laxminarayan & Macauley (2012).

English, L. (1999). Improving Data Warehouse and Business Information Quality: Methods for reducing costs and increasing Profits. New York, NY, USA, Willey & Sons.

English, L. P. (2009) Information Quality Applied: Best Practices for Improving Business Information Processes. Wiley Publishing.

Eppler, M. (2001). A Generic Framework for Information Quality in Knowledge-Intensive Processes. Proceeding of the Sixth International Conference on Information Quality.

Eppler, M. y Helfert, M. (2004). A Classification and Analysis of Data Quality Costs. International Conference on Information Quality, MIT, Cambridge, MA, USA.

Even, A. y Shankaranarayanan, G. (2007). Utility-driven assessment of data quality. ACM SIGMIS Database, 38(2), 75-93.

FFIEC (2017). Cybersecurity Assesment Tool. Federal Financial Institutions Examination Council, Mayo 2017.

Fishbein, M. y Ajzen, I. (1975). Beliefs, Attitude, Intention and Behavior: An Introduction to Theory and Research, Addison-Wesley, Reading, MA.

Fryman, L., Lampshire, G. y Meers, D. (2016). The Data and Analytics Playbook: Proven Methods for Governed Data and Analytic Quality. Morgan Kaufmann.

García, F., Bertoa, M., Calero, C., Vallecillo, A., Ruiz, F., Piattini, M. y Genero, M (2006). Towards a consistent terminology for software measurement. Information & Software Technology 48(8): 631-644.

Gartner (2017). Master Data Management (MDM). Recuperado 10 de noviembre de 2017, a partir de https://www.gartner.com/it-glossary/master-data-management-mdm.

Galdies, P. (2014). Business Intelligence. The insider threat to data assets. Journal of Direct, Data and Digital Marketing Practice, Vol. 15, Nº 3, 197-200.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:33:09.

Guerra-García, C., Caballero, I. y Piattini, M. (2012). "Capturing data quality requirements for web applications by means of DQ_WebRE." Information Systems Frontiers: 1-13.

Harari, Y.N. (2016). Homo Deus. A Brief History of Tomorrow. Reino Unido, Penguin Random House.

Heinrich, B. y Klier. M. (2015). Metric-based data quality assessment – Developing and evaluating a probability-based currency metric. Decision Support Systems 72, April, 82-96.

Heinrich, B. y Hristova, D. (2016). A quantitative approach for modelling the influence of currency of information on decision-making under uncertainty. Journal of Decision Systems 25 (1), 16-41.

Helfert, M. y Zakir, F. M. (2010). An approach to monitoring data quality – product oriented approach. Proc. of the Sixteenth Americas Conference on Information Systems, Agosto, 2010, 1-11.

Hevner, A.R., S.T. March, J. Park, y S. Ram (2004). Design Science in Information Systems Research. MIS Quarterly 28 (1): 75-105.

ISACA. (2012). COBIT 5: A Business Framework for the Governance and Management of Enterprise IT: ISACA.

ISACA (2013). COBIT 5 Información Catalizadora. ISACA, Rolling Meadows, IL, EE.UU.

ISO. (2008). ISO 2709 Information and documentation: Format for information exchange. Ginebra: International Organization for Standarization.

ISO. (2009a). ISO 8000-102 Master Data: Exchange of characteristic data: Vocabulary. Ginebra: International Organization for Standarization.

ISO (2009b). ISO 8000-110 Data quality – Part 110: Master data: Exchange of characteristic data: Syntax, semantic encoding, and conformance to data specification. Ginebra: International Organization for Standardization.

ISO (2011). ISO/TS 8000-150 Data quality – Part 150: Master data: Quality management framework. Ginebra, International Organization for Standardization.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:33:09.

ISO (2012). ISO/TS 19158 Geographic Information – Quality assurance of data supply. Ginebra: International Organization for Standardization.

ISO (2013). ISO 19157 Geographic Information – Data Quality. Ginebra: International Organization for Standardization.

ISO (2015). ISO 8000-8 Data quality – Part 8: Information and data quality: Concepts and measuring. Ginebra: International Organization for Standarization.

ISO (2016a). ISO 8000-100 Data quality – Part 100: Master data: Exchange of characteristic data: Overview. Ginebra: International Organization for Standardization.

ISO (2016b). ISO 8000-120 Data quality – Part 120: Master data: Exchange of characteristic data: Provenance Ginebra: International Organization for Standardization.

ISO (2016c). ISO 8000-130 Data quality – Part 130: Master data: Exchange of characteristic data: Accuracy. Ginebra: International Organization for Standardization.

ISO (2016d). ISO 8000-140 Data quality – Part 140: Master data: Exchange of characteristic data: Completeness. Ginebra: International Organization for Standardization.

ISO (2016e). ISO 8000-61 Data quality – Part 61: Data quality management: Process reference model. Ginebra: International Organization for Standardization.

ISO (2018a). ISO 8000-115 Master data: Exchange of quality identifiers: Syntactic, semantic and resolution requirements. Ginebra: International Organization for Standardization.

ISO (2018b). ISO 8000-116 Data quality: Application of ISO 8000-115 to the formatting of Authoritative Legal Entity Identifiers (ALEI) for individuals and organizations. Ginebra: International Organization for Standardization.

ISO/IEC (2009). ISO/IEC 25012 Software engineering. Software product quality requirements and evaluation (SQuaRE). Data quality model. Ginebra: International Organization for Standarization.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:33:09.

ISO/IEC (2015a). ISO/IEC 25024 Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Measurement of data quality. Ginebra: International Organization for Standarization.

ISO/IEC (2015b). ISO/IEC 33002 Information technology – Process assessment – Requirements for performing process assessment.

ISO/IEC (2015c). ISO/IEC 33010 Information technology: Process assessment: Guide to performing assessment.

ISO/IEC (2015d). ISO/IEC 33020 Information technology: Process assessment: Process measurement framework for assessment of process capability.

ISO/IEC (2017) ISO/IEC 38505-1 – Information Technology – Governance of IT – Part 1: The application of ISO/IEC 38500 to the governance of data. Ginebra, Suiza.

ISO/IEC (2018) ISO/IEC TR 38505-2, Governance of IT – Part 2: Implications of 38505-1 for data management. Ginebra, Suiza.

Jochen, K. y Weisbecker, A. (2014). Master Data Management: Products and Research. Recuperado a partir de https://mafiadoc.com/master-data-management-citeseerx_599e09781723dd0a40e05e25.html.

Kassaa, J.M., González, J., Cuevas, A., Cuevas, R., Marciel, M. y Gonzalez, R. (2016). Your Data in the Eyes of the Beholders. Design of a unified data valuation portal to estimate value of personal information from market perspective. 11th International Conference on Availability, Reliability and Security.

Kenett, R.S. y Shmueli, G. (2017). Information Quality. The Potential of Data and Analytics to Generate Knowledge. UK, Wiley.

Kontio, M., Marttila-Kontio, M. y Hotti, V. (2015). Data Assessment Model for Strategic Management", Communications of the IBIMA, Vol. 2015 (2015), Article ID 561648, DOI:10.5171/2015.561648.

Laney, D.B. (2018). Infonomics. How to monetize, manage, and measure information as an asset for competitive advantage. Nueva York, EE.UU. Bibliomotion.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:33:09.

Laxminarayan, R. y Macauley, M.K. (2012) (eds.). The Value of Information. Methodological Frontiers and New Applications in Environment and Health. Springer.

López, J.J. y Zarza, G. (2017). La Ingeniería del Big Data. Cómo trabajar con datos. Barcelona, Editorial UOC.

Loshin, D. (2009). Master Data Management. Morgan Kaufmann.

Loshin, D. (2010). The practitioner's guide to data quality improvement. Elsevier.

Loshin, D. (2013). Business Intelligence, Morgan Kauffman.

Loucks, J., Macaulay, J., Noronha, A. y Wade, M. (2016). Digital Vortex. How Today´s Market Leaders Can Beat Disruptive Competitors at Their Own Game. Global Center for Digital Business Transformation. Lausanne, Suiza.

Marr, B. (2017). Data Strategy: How to Profit from a World of Big Data, Analytics and the Internet of Things. Kogan Page Ltd., Londres, Reino Unido.

McGilvray, D. (2008). Executing Data Quality Project: Ten Steps to Quality Data and Trusted Information. Burlington, MA, USA, Morgan Kaufmann.

Moody, D. y Walsh, P. (1999). Measuring the value of information: An asset valuation approach. Proceedings of the 7th European Conference on Information Systems, Denmark.

Caballero, I. C. (2004). Modelo para la evaluación y mejora de la gestión de la calidad de los datos y de la información. Tesis doctoral. Universidad de Castilla-La Mancha.

Nelson, H.J, Poels, G., Genero, M. y Piattini, M. (2012). A conceptual modeling quality framework. Software Quality Journal 20(1), 201-228.

NIST (2017). Framework for Improving Critical Infrastructure Cybersecurity. Draft Version 1.1, National Institute of Standards and Technology, January 2017.

Ortiz, M.A. (2006). Intellectual Capital (Intangible Assets) Valuation Considering The Context. Journal of Business & Economics Research 4 (9).

Otto, B. (2015). Quality and Value of the Data Resource in Large Enterprises. Information Systems Management 32, 234-251.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:33:09.

Pearce, G. (2017). Boosting Cyber Security With Data Governance and Enterprise Data Management. ISACA Journal Vol. 3, 1-6.

Peffers, K., T. Tuunanen, M.A. Rothenberger, and S. Chatterjee (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems* 24 (3): 45-77. doi:10.2753/MIS0742-1222240302.

Piattini, M. y Díaz, O. (eds.) (2000). Advanced Databases Technology and Design. UK, Artech House.

Piattini, M., Calero, C. y Genero, M. (2002). Information and database quality. Kluwer Academic Publishers, Norwell, USA.

Pries-Heje, J., y Johansen, J. (2013). ImprovAbility-Success with Process Improvement. DELTA.

Ramírez, A. (2017). Digitalízate o desaparece. Claves para transformarse y competir en la nueva era. Barcelona, Gestión 2000.

Raskino, M. y Waller, G. (2015). Digital to the Core. Oxon, Bibliomotion, UK.

Redman, T. C. (1996). Data Quality for the Information Age. Boston, MA, USA, Artech House Publishers.

Redman, T. C. (2008). Data driven: profiting from your most important business asset, Harvard Business School Press.

Repo, A. J. (1989). The Value of lnformation: Approaches in Economics, Accounting, and Management Science. Journal of The American Society for Information Science. 40(2), 68-85.

Roger, D.L. (2016). The Digital Transformation Playbook. USA, Columbia Business School.

Ross, A. (2016). The Industries of the Future. London, Simon & Schuster.

Rotella, P. (2012). Is Data The New Oil? Forbes, April 2nd 2012. *https://www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/#4ca3851b7db3*.

Ruiz, F., Vizcaino, A., Piattini, M. y García, F. (2004). An Ontology For The Management Of Software Maintenance Projects. International Journal of Software Engineering and Knowledge Engineering 14(3), 323-349.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:33:09.

Runeson, P., Host, M., Rainer, A. y Regnell, B. (2012). Case Study Research in Software Engineering: Guidelines and Examples. Wiley.

Sarsfield, S. (2009). The Data Governance Imperative, IT Governance Publishing.

Saez-Trumper, D., Liu, Y., Baeza-Yates, R., Krishnamurthy, B. y Mislove, A. (2014). Beyond CPM and CPC: determining the value of users on OSNs. ACM COSN 14

SAS (2017). Value Information as an Asset. White Paper. Accedido en diciembre 2017. Disponible en: *http://www.eurim.org.uk/activities/ig/InformationAsset.pdf*

Schmarzo, B. y Sidaoui, M. (2017). Applying economic concepts to big data to determine the financial value of the organization's data and analytics, and understanding the ramifications on the organizations' financial statements and IT operations and business strategies.EMC. *https://infocus.emc.com/wp-content/uploads/2017/04/USF_The_ Economics_*of_Data_and_ Analytics-Final3.pdf

SEI (2014). DMM: Data Management Maturity Model. Software Engineering Institute Pittsburgh.

Shen, Y., Guo, B., Shen, Y., Duan, X., Dong, X., y Zhang, H. (2016). A Pricing Model for Big Personal Data. Tsinghua Science and Technology 21 (5), 482-490.

Short, J.E. y Todd, S. (2017). What´s Your Data Worth?. MIT Sloan Management Review, 58 (3), 16-19.

Shull, F.; Singer, J. y Sjøberg, D. I.K. (eds.) (2008). Guide to Advanced Empirical Software Engineering. Springer.

Sidgman, J. y Crompton, M. (2016). Valuing Personal Data to Foster Privacy: A Thought Experiment and Opportunities for Research. Journal of Information Systems American Accounting Association 30 (2), 169–181.

Siyam, G. I., Wynn, D.C., y Clarkson, P.J. (2015). Review of Value and Lean in Complex Product Development. Systems Engineering 18 (2), 192–207.

Soares, S. (2010). The IBM Data Governance Unified Process: Driving Business Value with IBM Software and Best Practices: MC Press, LLC.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:33:09.

Spiekermann, S. y Korunovska, J. (2017). Towards a Value Theory for Personal Data. Journal of Information Technology. 32 (1), 62-84.

Strong, D. M., et al. (1997). "Data Quality in Context." Communications of the ACM 40(5): 103-110.

Sun, H., Fang, Y. y Hsieh, J.J. (2014). Consuming information systems: An economic model of user satisfaction. Decision Support Systems 57, 188–199.

Talburt, J. R. (2011). Entity resolution and information quality. Elsevier.

Terzi, D. S., Terzi, R., y Sagiroglu, S. (2015), A Survey on Security and Privacy Issues in Big Data. The 10th International Conference for Internet Technology and Secured Transactions (ICITST-2015). IEEE Computer Society, 202-208.

Tupper, C. (2011). Data architecture: from zen to reality. Elsevier.

Vizcaíno, A., García, F., Piattini, M. y Beecham, S. (2016). A validated ontology for global software development. Computer Standards & Interfaces 46, 66-78.

Vu, Q.H., Pham, T, Truong, H., Dustdar, S. y Asal, R. (2012). DEMODS: A Description Model for Data-as-a-Service. 26th IEEE International Conference on Advanced Information Networking and Applications.

Walter, P., Werth, D., y Loos, P. (2006). Peer-to-Peer-Based Model-Management for Cross-Organizational Business. En 15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE'06) (pp. 255-260). *https://doi.org/10.1109/WETICE.2006.53*

Wand, Y. y Wang, R. Y. (1996). Anchoring Data Quality Dimensions in Ontological Foundations. Communications of the ACM 39(11): 86-95.

Wang, R. Y. (1998). A product perspective on total data quality management. Communications of the ACM, 41(2), 58–65.

Wang, R. Y., Lee, Y. W., Pipino, L. L., y Strong, D. M. (1998). Manage your information as a product. Sloan Management Review, 39(4), 95–105.

Westerman, G., Bonnet, D., y McAfee, A. (2014). Leading Digital. Turning Technology into Business Transformation. USA, Harvard Business Review.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:33:09.

Wieringa, R. (2009). Design Science as Nested Problem Solving. Paper presented at the 4th International Conference on Design Science Research in Information Systems and Technology, Philadelphia, May 7-8.

Wohlin, C., Runeson, P., Höst. M., Ohlson, M., Regnell, B., Wesslén, A. (2012). Experimentation in Software Engineering: An Introduction. Springer.

Yin, R. K. (2014). Case Study Research. 5ª ed. SAGE Publications.

Zechmann, A. (2016). Assessing the Economic Value of Data Assets. Work Report. University of St. Gallen.

Copyright © 2018. RA-MA Editorial. All rights reserved.

Gualo, Cejudo, Fernando, et al. Calidad de datos, RA-MA Editorial, 2018. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/bibudlimasp/detail.action?docID=5758915.
Created from bibudlimasp on 2020-02-28 07:33:09.