Gaona, G., Bollati, V. A., Caballero Muñoz-Reja, I., Gualo, F.: Diseño de un sistema de gobierno de datos: un marco de trabajo para la priorización de procesos. In: Durán Toro, A. (ed.) Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2023). Sistedes (2023). https://hdl.handle.net/11705/JISBD/2023/6178

Diseño de un sistema de gobierno de datos: un marco de trabajo para la priorización de procesos

Germán Gaona[1[0000-0002-4148-705X]], Verónica A. Bollati[2[0000-0003-4008-9903]], Ismael Caballero[3[0000-0002-5189-1427]], Fernando Gualo[3,4[0000-0002-7800-7902]]

[1] UTN - FRRE, Resistencia, Argentina
gg_rcia@ca.frre.utn.edu.ar
[2] CONICET - UTN - FRRE, Resistencia, Argentina
vbollati@ca.frre.utn.edu.ar
[3] Instituto de Tecnologías y Sistemas de Información (Alarcos Research Group), UCLM, Ciudad Real, Spain
Ismael.Caballero@uclm.es
[4] DQTeam, Ciudad Real, Spain
fgualo@dqteam.es, Fernando.Gualo@uclm.es

Abstract. Las organizaciones necesitan gobernar sus datos para asegurar que se consigan los objetivos establecidos en sus estrategias organizacionales de datos. No obstante, el gobierno de datos también debe servir para resolver o evitar problemas, cuya manifestación suele darse bajo la forma de niveles inadecuados de calidad para los datos usados en ciertos contextos. Una de las consecuencias más interesantes del gobierno de datos es la implantación de procesos de gestión de datos, gestión de calidad de datos y gobierno de datos para que la organización pueda abordar de forma sistemática y rigurosa estos problemas de calidad de datos. El Modelo Alarcos de Madurez de Datos (MAMD) proporciona la definición de estos procesos. En este artículo se presenta un Marco para la Priorización de Procesos de MAMD (MPPM) con el objetivo de guiar a las organizaciones en la identificación de qué procesos de MAMD se deberían implementar para evitar dichos problemas, partiendo de los obstáculos de calidad de datos más comunes en las organizaciones. Como prueba de concepto, este marco de trabajo se aplicó en el contexto de la Universidad Tecnológica Nacional (Argentina).

Keywords: Gobierno de datos, MAMD, marco de priorización de procesos

1 Introducción

Un proceso de toma de decisiones eficaz es crucial para el éxito de las organizaciones, y dicha eficacia está estrechamente relacionada con la calidad de los datos utilizados. Sin embargo, el volumen y la diversidad de los datos involucrados, son sólo algunos de los desafíos a los que las organizaciones se enfrentan en la actualidad [1].

En el ámbito del gobierno de las organizaciones [2], se promueve el reconocimiento de los datos como un activo valioso para ellas, siendo el desarrollo de una estrategia de gobierno de datos, en particular, esencial para el logro de este propósito [3].

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.

Gaona, G., Bollati, V. A., Caballero Muñoz-Reja, I., Gualo, F.: Diseño de un sistema de gobierno de datos: un marco de trabajo para la priorización de procesos. In: Durán Toro, A. (ed.) Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2023). Sistedes (2023). https://hdl.handle.net/11705/JISBD/2023/6178

Un sistema de gobierno de datos (SGD) efectivo permite alinear los datos con los objetivos estratégicos [4] y establecer las responsabilidades acerca de la gestión de datos y la gestión de la calidad de los datos a nivel organizativo.

Según COBIT 2019 [5], un sistema de gobierno tiene siete componentes, siendo uno de ellos la colección de procesos a seguir. En el contexto de un SGD estos procesos deberían vincularse con la gestión de datos, la gestión de la calidad de datos y el gobierno de datos.

Existen varios marcos de trabajo de gobierno y gestión de datos que incluyen una descripción de procesos que pueden ser tomados como referencia. En este trabajo se tomará como referencia el Modelo Alarcos de Madurez de Datos (MAMD), ya que de acuerdo a [6] y [7] posee ciertas ventajas, como su gratuidad, genericidad y abordaje holístico del SGD.

Dado el carácter estratégico del gobierno del dato, un SGD no debe implantarse de manera abrupta, ni del mismo modo en todos los ámbitos, ya que este es afectado por diversos factores de contingencia [8]; considerando la cantidad de procesos definidos en MAMD, es necesario establecer criterios de selección y priorización de dichos procesos de acuerdo con la realidad de cada organización.

Considerando lo anterior, la motivación de esta investigación es proporcionar a los analistas de negocio y a los analistas de calidad de datos un marco de trabajo que permita identificar y priorizar los procesos que deberían implementarse como parte del SGD. Específicamente, el marco de trabajo presentado se denomina Marco para la Priorización de Procesos de MAMD (MPPM) y consta de:

1. Una tipología de causas raíz potenciales relacionadas con niveles inadecuados de calidad de datos, obtenidas de [9]. Éstas serán útiles para el diagnóstico de problemas más importantes.
2. La colección de procesos de MAMD vinculados con la gestión de datos, la gestión de la calidad de datos y el gobierno de datos.
3. Un método para la clasificación de los problemas específicos encontrados en las organizaciones en las causas raíz probables.
4. Un método para la selección y priorización de los procesos del modelo de referencia de MAMD a implantar, basados en el tipo y grado de contribución de dichos procesos a la solución de los problemas detectados.

MPPM fue evaluado en el dominio de la Universidad Tecnológica Nacional - Facultad Regional Resistencia (UTN-FRRe). La motivación de esto radica en las ventajas que traería la implantación de un SGD, considerando las recomendaciones de mejora en la gestión de la información efectuadas por la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU) [10].

El resto del artículo está organizado de la siguiente manera: en la sección 2, se muestran los aspectos más relevantes de los conceptos necesarios para comprender el trabajo; en la sección 3 se presenta la propuesta del marco de trabajo; en la sección 4 se describe la validación del marco en el caso de una universidad argentina. Por último, en la sección 5, se presentan las conclusiones y trabajos futuros.

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.

Gaona, G., Bollati, V. A., Caballero Muñoz-Reja, I., Gualo, F.: Diseño de un sistema de gobierno de datos: un marco de trabajo para la priorización de procesos. In: Durán Toro, A. (ed.) Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2023). Sistedes (2023). https://hdl.handle.net/11705/JISBD/2023/6178

2 Conceptos previos

En esta sección se presentan algunos conceptos necesarios para el correcto entendimiento de la propuesta.

2.1 Problemas de datos y sus causas raíz

En la mayoría de las organizaciones existen problemas derivados de niveles inadecuados de calidad de datos que perjudican el desempeño de éstas en sus distintos niveles, por lo que es importante detectarlos y corregirlos a tiempo. Para proveer una solución adecuada, es fundamental descubrir de forma precisa las causas raíz que los originan, y determinar cómo abordar dichos problemas desde la dirección impuesta por el gobierno de datos.

En el estudio presentado en [9] se describen diez tipos de causas raíz vinculadas con la calidad de datos, obtenidas de tres organizaciones diferentes. Las causas raíz codificadas se enumeran a continuación:

- (CR1) *"Múltiples fuentes de la misma información producen diferentes valores."*
- (CR2) *"La información se produce utilizando juicios subjetivos, lo que lleva a un sesgo."*
- (CR3) *"Los errores sistémicos en la producción de información conducen a la pérdida de información."*
- (CR4) *"Los grandes volúmenes de información almacenada dificultan el acceso a la información en un tiempo razonable."*
- (CR5) *"Los sistemas heterogéneos distribuidos conducen a definiciones, formatos y valores inconsistentes."*
- (CR6) *"La información no numérica es difícil de indexar."*
- (CR7) *"El análisis de contenido automatizado en todas las colecciones de información aún no está disponible."*
- (CR8) *"Cambios en las necesidades de información de los consumidores."*
- (CR9) *"El fácil acceso a la información puede entrar en conflicto con los requisitos de seguridad, privacidad y confidencialidad."*
- (CR10) *"La falta de recursos informáticos suficientes limita el acceso."*

Este catálogo de tipos de causas raíz puede ser aplicado como una herramienta efectiva para identificar y diagnosticar las causas subyacentes sobre problemas específicos en diferentes dominios y explorar alternativas no contempladas. Además, esta tipología de problemas de causas raíz sirve también para organizar, completar y simplificar la información relacionada con el problema en base al conjunto de categorías establecidas, tal como se plantea en otros enfoques [11].

2.2 Gobierno de datos y marcos de referencia

Los problemas de calidad de datos y sus causas raíz referidos anteriormente, no suelen ocurrir de manera aislada. Para poder proveer una solución eficaz no alcanza con la

This work is licensed under a Creative Commons "Attribution-NonCommercial-NoDerivatives 4.0 International" license.

Gaona, G., Bollati, V. A., Caballero Muñoz-Reja, I., Gualo, F.: Diseño de un sistema de gobierno de datos: un marco de trabajo para la priorización de procesos. In: Durán Toro, A. (ed.) Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2023). Sistedes (2023). https://hdl.handle.net/11705/JISBD/2023/6178

mera gestión de los datos y de su calidad, también es necesario realizar un abordaje desde la perspectiva del gobierno de datos que ayude a instaurar estrategias, políticas y directrices claras a nivel organizacional.

El gobierno de datos es "*el ejercicio de autoridad y control (planificación, monitoreo y aplicación) sobre la gestión de los activos de datos*" [12].

Los modelos o marcos de referencias existentes permiten guiar la implantación de un SGD y definir un camino de mejora; la mayoría de ellos establecen una cierta cantidad de niveles de madurez y dominios a los que apuntan.

Algunos de estos modelos son: *IBM Data Governance Maturity Model* [13], *DAMA-DMBOK Framework* [12], *EDM Council's DCAM Framework* [14] y MAMD [15], cuyo modelo de referencia de procesos está constituido por los procesos recogidos en las especificaciones *UNE 0077: Gobierno del dato* [5], *UNE 0078: Gestión del dato* [16] y *UNE 0079: Gestión de la calidad del dato* [17]. La especificación [5] establece una distinción clara entre las disciplinas de gobierno del dato y gestión del dato, definiendo los procesos y áreas de actuación asociadas a cada una de ellas.

Cada proceso de MAMD es especificado por su nombre, un propósito, los resultados de procesos esperados, las tareas para conseguir estos resultados y el conjunto de productos de trabajos involucrados.

3 Propuesta

Tal como se planteó en la introducción, el propósito de MPPM es identificar y priorizar aquellos procesos de MAMD que se deben implantar considerando su aportación a la solución de los problemas específicos detectados en una organización particular.

MPPM consta de cuatro partes: (1) un catálogo de tipos de causas raíz vinculadas a problemas de calidad de datos, (2) la colección de procesos de MAMD, (3) un método para la clasificación de los problemas específicos a dichas causas raíz y (4) un método para la selección de los procesos prioritarios a implantar. En el resto de la sección se explica cada parte en detalle.

3.1 Un catálogo de tipos de causas raíz potenciales

La priorización de procesos descrita para MPPM requiere analizar la naturaleza de los problemas de datos encontrados en cada organización.

El compendio presentado en [9] (véase sección 2.1 y sección 3.3) es utilizado aquí como referencia de las causas raíz más comunes, convirtiéndolo en una herramienta eficaz para estudiar y clasificar los problemas específicos encontrados, además de explorar alternativas no contempladas.

De este compendio se desprenden dos características deseables:

- Cada causa raíz es descrita de manera sistemática incluyendo ejemplos, dimensiones o características de calidad de datos afectadas, efectos organizacionales, señales de alerta, soluciones temporales, los problemas con ellas y las soluciones reales propuestas.

This work is licensed under a Creative Commons "Attribution-NonCommercial-NoDerivatives 4.0 International" license.

Gaona, G., Bollati, V. A., Caballero Muñoz-Reja, I., Gualo, F.: Diseño de un sistema de gobierno de datos: un marco de trabajo para la priorización de procesos. In: Durán Toro, A. (ed.) Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2023). Sistedes (2023). https://hdl.handle.net/11705/JISBD/2023/6178

- El nivel de abstracción con que se enuncian las causas raíz: esto constituye una de las razones de su vigencia a pesar de su relativa antigüedad y le otorga suficiente flexibilidad para ser aplicado a distintas situaciones.

3.2 La colección de procesos de MAMD

MAMD [15] (y sus 23 procesos recogidos en las especificaciones [5, 16, 17]) es el marco de referencia elegido para esta propuesta ya que permite guiar la implantación de un SGD y definir un camino de mejora de acuerdo con un modelo de madurez.

Este modelo de madurez aporta una idea de qué procesos serían prioritarios para la implantación de un SGD, sin embargo, es necesario realizar una evaluación particular para cada organización de acuerdo con su aportación de dichos procesos al problema de datos a resolver. Para el desarrollo de MPPM se adoptan los procesos de MAMD por los siguientes motivos:

- El modelo de referencia de proceso es abierto.
- Posee una clara separación de los procesos vinculados a la gestión de datos, la gestión de calidad de datos y el gobierno de datos, pero a su vez coordinados entre sí.
- La descripción de cada proceso tiene el nivel de abstracción suficiente para ser manejable y adaptable a diferentes contextos. Cada proceso tiene un nombre, un propósito, los resultados de procesos esperados, las tareas para conseguir estos resultados y el conjunto de productos de trabajos involucrados.
- Es certificable según el esquema de las normas ISO/IEC 33000.

3.3 Método para la clasificación de los problemas específicos

Un elemento importante de MPPM es la asignación de los problemas específicos hallados en una organización a las causas raíz genéricas probables. Este método consta de dos partes: (1) criterios para la clasificación y (2) una matriz que relaciona y agrega los problemas específicos con las causas raíz genéricas.

Esta clasificación previa es requerida pues la priorización de procesos se efectúa en base a las causas raíz genéricas, de esta manera se logra que el marco de trabajo sea aplicable a diferentes dominios independientemente de los problemas particulares.

(1) Criterios de clasificación. Para realizar la clasificación se proponen dos criterios inspirados en las etapas y las señales de alerta vinculadas a las causas raíz descritas en [9]. Las aportaciones de este elemento de MPPM se centran en el refinamiento de las señales de alerta descritas en la propuesta original y su utilización como un criterio para la clasificación. Los criterios descritos a continuación se presentan en el orden sugerido para su evaluación:

1. *Detección del problema:* el propósito de este criterio es reducir el espectro de causas raíz probables de acuerdo con la etapa en la que se produce el problema específico. Se consideran tres etapas, aquella en la que se producen datos (producción), aquella en la que se almacenan, mantienen o aseguran datos (almacenamiento), y, por último, aquella en la que se los utilizan (utilización).

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.

Gaona, G., Bollati, V. A., Caballero Muñoz-Reja, I., Gualo, F.: Diseño de un sistema de gobierno de datos: un marco de trabajo para la priorización de procesos. In: Durán Toro, A. (ed.) Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2023). Sistedes (2023). https://hdl.handle.net/11705/JISBD/2023/6178

2. *Señales de alerta del problema:* una vez determinada la etapa en la que ocurre el problema específico, se buscan entre las señales de alerta que se encuentran vinculadas con esa etapa. Para este trabajo se enriquecieron las señales de alerta existentes en el estudio original, aquí estas se describen como un conjunto de indicios que ayudan a determinar la causa raíz probable para cada problema específico.

En la Tabla 1 se presentan los criterios descritos, para cada etapa se detallan las señales de alerta asociadas. Cuando un problema específico cumple con la mayoría de las señales de alerta asociadas con una causa raíz es probable que dicho problema deba vincularse con dicha causa.

Tabla 1. Criterios para asignación de problemas específicos a causas raíz.

Etapa	C. raíz	Señales de alerta
Producción	CR1	Se presentan varios procesos que generan el mismo elemento de datos de forma independiente. Varios sistemas de información manejan el mismo elemento de dato. Los valores obtenidos de diversas fuentes de datos no concuerdan. La inconsistencia de los datos que proporciona el sistema de información hace que las personas desconfíen de ellos.
	CR2	Existen actividades que generan datos a través de juicios subjetivos. Dichas actividades no poseen reglas claras para la generación de esos datos. No existen sistemas expertos que brinden soporte para generar esos datos.
	CR3	Se detectan errores repetitivos en los mismos elementos de datos. Las validaciones en formularios de entrada de datos son inadecuadas Incentivos inadecuados para las personas que cargan los datos.
Almacenamiento	CR4	Existen grandes volúmenes de datos almacenados que deben ser accedidos para el trabajo diario. El tiempo o la velocidad de acceso son excesivos.
	CR5	Existen varios sistemas de información en la organización. Los datos son difíciles de consolidar debido a inconsistencias en el formato. Los valores para el mismo elemento de dato son inconsistentes. Estos datos se almacenan en repositorios diferentes que requieren duplicación de esfuerzo para su mantenimiento.
	CR6	Existen datos almacenados en formato de audio, imagen, video, documentos o texto no estructurado. Existen dificultades para acceder de manera eficiente a esos datos. El beneficio de su almacenamiento es bajo o no está claramente definido.
Utilización	CR7	Además de las señales enunciadas para CR6, se agregan: Capacidades limitadas de manipulación de los datos no estructurados almacenados. Capacidades analíticas limitadas sobre los datos no estructurados almacenados. Estos datos se encuentran almacenados en múltiples sistemas con diferentes nombres, definiciones y formatos.
	CR8	Existen diferencias entre los datos provistos a los consumidores no se ajustan a sus necesidades.

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.

	Estas diferencias pueden deberse a cambios en sus preferencias, en las tareas que realizan, en el entorno organizacional entre otros.
CR9	Se presentan dificultades de acceso a los datos relacionados con los requisitos de seguridad, privacidad y confidencialidad. Estas dificultades incluyen: barreras de acceso a los datos, retrasos en el acceso a los datos, falta de datos relevantes, ausencia de colaboración y pérdida de productividad.
CR10	Existen quejas reiteradas comprobables de los usuarios acerca del funcionamiento del sistema informático en cuanto a: su lentitud, la disponibilidad de dispositivos, las limitaciones de almacenamiento, fallos frecuentes y acceso deficiente a la red, entre otros.

(2) Matriz de vinculación de problemas específicos y causas raíz asociadas. La matriz de vinculación tiene como propósito: (1) registrar cada problema específico detectado, (2) indicar su asignación a una o más causas raíz genéricas y (3) contabilizar el total de asignaciones a cada causa, lo que sirve de entrada para el procedimiento de priorización final.

En la Tabla 2 se ilustra de forma genérica y resumida la matriz de vinculación para *p* problemas específicos del dominio bajo estudio y las diez causas raíz tomadas de referencia. Dónde, los problemas específicos (PE) se indican en cada fila y las causas raíz (CR), en cada columna; en la intersección de ambos se especifica si existe asignación (X). Antes de finalizar el procedimiento se contabilizan (en la última fila) el total de problemas específicos asignados a cada causa raíz (TCR).

Tabla 2. Matriz de vinculación genérica

	CR1	CR2	…	CR10
PE1		X	…	
PE2			…	X
…	…	…	…	…
PEp			…	X
Cantidad	TCR1	TCR2	…	TCR10

3.4 Selección de los procesos prioritarios a implantar

El objetivo principal de MPPM radica en la recomendación de aquellos procesos de MAMD a implantar que son considerados prioritarios de acuerdo con su aporte a la solución de las causas raíz.

Este componente se conforma por dos partes: (1) clasificación de los procesos de MAMD según su contribución a la solución a las causas raíz y (2) obtención de los procesos prioritarios a implantar a través de una matriz de priorización. A continuación, se detallarán cada una de las partes:

(1) Clasificación de los procesos de MAMD según su contribución a la solución. Para efectuar la recomendación de los procesos a implantar, en primer lugar, ha sido necesario realizar la clasificación de cada proceso de MAMD respecto de su contribución a la solución de cada causa raíz probable. Como esta clasificación no depende de los problemas específicos, puede ser reutilizado en diferentes dominios.

Gaona, G., Bollati, V. A., Caballero Muñoz-Reja, I., Gualo, F.: Diseño de un sistema de gobierno de datos: un marco de trabajo para la priorización de procesos. In: Durán Toro, A. (ed.) Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2023). Sistedes (2023). https://hdl.handle.net/11705/JISBD/2023/6178

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.

Gaona, G., Bollati, V. A., Caballero Muñoz-Reja, I., Gualo, F.: Diseño de un sistema de gobierno de datos: un marco de trabajo para la priorización de procesos. In: Durán Toro, A. (ed.) Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2023). Sistedes (2023). https://hdl.handle.net/11705/JISBD/2023/6178

La técnica de clasificación consta de cuatro elementos: (1) un grupo de expertos, (2) los criterios de clasificación, (3) el procedimiento de clasificación y (4) la matriz de clasificación resultante. A continuación, se explican cada uno de ellos.

Grupo de expertos. La clasificación se basa en la opinión de cada integrante de un grupo de expertos de una o más disciplinas vinculadas con la gestión de datos, gestión de calidad de datos o gobierno de datos. El número de integrantes debe ser preferente mayor a dos e impar, pues podría ser necesario dirimir discrepancias en la clasificación.

Criterios de clasificación. La clasificación debe ser realizada según dos criterios: el tipo de aportación y la importancia de dicha aportación del proceso a la solución de la causa raíz.

- *Tipo de aportación:* indica la naturaleza de la aportación del proceso a la solución de la causa raíz. Esta clasificación se basa en la descrita en [18]:
  - Operativo (O): indica que la implantación del proceso contribuye de manera directa a la solución del problema cuya causa raíz se ha identificado.
  - Soporte (S): indica que la implantación del proceso contribuye de manera indirecta a la solución de la causa raíz actuando sobre su entorno.
  - Estratégico (E): indica que la implantación del proceso contribuye de manera indirecta a la solución de la causa raíz actuando a nivel organizacional.
- *Importancia de la aportación:* indica el grado de importancia de la implantación de determinado proceso de MAMD con respecto a la solución de la causa raíz analizada. Esta clasificación se basa en el método de priorización MoSCoW [19]:
  - *Must have* (M): son procesos obligatorios con un alto impacto en la solución de la causa raíz.
  - *Should have* (S): son procesos altamente deseables con un impacto medio en la solución de la causa raíz.
  - *Could have* (C): son procesos deseables con un impacto bajo en la solución de la causa raíz.
  - *Will not have* (W): son procesos mínimamente deseables con un impacto nulo o casi nulo en la solución de la causa raíz.

Procedimiento de clasificación. Aquí se indica cómo se lleva a cabo la clasificación. El procedimiento debe contar al menos con las siguientes actividades:

- Reunir a los expertos y explicar claramente el objetivo de la tarea, los criterios a utilizar, el procedimiento de clasificación, las causas raíz y los procesos de MAMD. Se debe entregar todo el material necesario para la ejecución de la tarea (p.ej. documentación aclaratoria, especificaciones, etc.).
- Cada experto se encarga de realizar la clasificación de manera separada, considerando los criterios mencionados, pudiendo justificar de manera escrita sus opiniones.
- Integrar las diferentes clasificaciones individuales realizadas por los expertos, en los casos que existan discrepancias, se trata de llegar a un consenso.

Matriz de clasificación. Los resultados tanto individuales como integrados se registran en una matriz de clasificación, en la que cada fila representa un proceso de los 22 existentes en MAMD (PM) y cada columna refleja una causa raíz de las 10 enumeradas

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.

Gaona, G., Bollati, V. A., Caballero Muñoz-Reja, I., Gualo, F.: Diseño de un sistema de gobierno de datos: un marco de trabajo para la priorización de procesos. In: Durán Toro, A. (ed.) Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2023). Sistedes (2023). https://hdl.handle.net/11705/JISBD/2023/6178

(CR). En la intersección de cada fila y columna se especifica un par ordenado <t; i>, dónde *t* es un valor de la escala según el tipo de aportación e *i* es un valor de la escala según la importancia de la aportación de cada proceso a la solución de determinada causa raíz. La matriz resultante se muestra en la Tabla 3.

Tabla 3. Matriz de clasificación genérica

	CR1	CR2	…	CR10
PM1	t; i	t; i	…	t; i
PM2	t; i	t; i	…	t; i
…	…	…	…	…
PM22	t; i	t; i	…	t; i

(2) Construcción de la matriz de priorización. Este componente genera la lista definitiva de procesos de MAMD con sus prioridades establecidas. Para ello se requiere como entrada: (1) la matriz de clasificación confeccionada con la opinión de los expertos acerca del aporte de los procesos a la solución y (2) la cantidad de problemas específicos asignados a cada causa raíz.

Para construir la *matriz de priorización*, a la matriz de clasificación descrita en el punto anterior, se le añaden: una fila con las cantidades de problemas específicos por causa raíz y una columna dónde se calculan y especifican las prioridades correspondientes a cada proceso.

El cálculo de la prioridad del proceso *i* (*PPi*), tal que $1 \leq i \leq$ m, se realiza según la ecuación 1:

$$PPi = \sum_{j=1}^{n} \left(a_{ij} b_{ij} c_j\right) \quad (1)$$

Siendo *m* la cantidad de procesos de MAMD a seleccionar y priorizar, *n* la cantidad total de causas raíz, a_{ij} es el valor asignado a la clasificación según el tipo de contribución, b_{ij} es el valor asignado a la clasificación según la importancia de la contribución y c_j es la cantidad de problemas específicos existentes por cada causa raíz.

Tabla 4. Matriz de priorización genérica final

	CR1	CR2	…	CR10	PP
Cantidad	TCR1	TCR2	…	TCR10	
PM1	t; i	t; i	…	t; i	PP1
PM2	t; i	t; i	…	t; i	PP2
…	…	…	…	…	…
PM22	t; i	t; i	…	t; i	PP22

Para facilitar el cálculo de la prioridad de cada proceso, se asignan valores numéricos a cada elemento de las escalas. A la escala del tipo de aportación se le asigna: O=3, S=2 y E=1; en tanto que, a la escala de la importancia de la aportación se le asigna: M=3, S=2, C=1 y W=0,5. La Tabla 4 muestra un ejemplo genérico de la matriz con los procesos priorizados.

Una vez calculada la prioridad para cada proceso, se deben seleccionar aquellos que posean el mayor valor de prioridad, el umbral de corte lo establece la persona usuaria del marco de trabajo de acuerdo con sus necesidades.

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.

Gaona, G., Bollati, V. A., Caballero Muñoz-Reja, I., Gualo, F.: Diseño de un sistema de gobierno de datos: un marco de trabajo para la priorización de procesos. In: Durán Toro, A. (ed.) Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2023). Sistedes (2023). https://hdl.handle.net/11705/JISBD/2023/6178

4 Aplicación

Como prueba de concepto de MPPM, en esta sección se presenta un ejemplo sintético en el dominio de la UTN-FRRe. La razón principal, tal como se mencionó en la introducción, radica en las ventajas asociadas con la implantación de un SGD y su relación con las evaluaciones periódicas y las recomendaciones realizadas por la CONEAU [10]. Por su parte, en [20] se plantea que una cultura *data-driven*, permitiría enfocarse en las áreas que tienen potencial de mejora en las universidades.

Para el relevamiento de los problemas específicos de calidad de datos en la UTN-FRRe, se efectuaron una serie de entrevistas con el Secretario Académico, el Director del Departamento de Alumnos, tres docentes y cuatro alumnos.

La descripción de este ejemplo se divide en tres partes, coincidentes con las contribuciones principales del trabajo: (1) la clasificación de los problemas de calidad de datos específicos hallados en el dominio de la UTN-FRRe vinculadas con las causas raíz genéricas, (2) la evaluación de los procesos de MAMD de acuerdo con su aporte a la solución de las causas raíz genéricas, y (3) la priorización de los procesos de MAMD considerando los dos puntos anteriores. Por las restricciones de espacio, para este ejemplo sólo se incluirán un subconjunto de las causas raíz y de los procesos de MAMD.

4.1 Clasificación de problemas específicos

A partir de las entrevistas realizadas se obtuvo un listado de problemas específicos de calidad de datos, para cada uno de ellos se asignaron las causas raíz correspondientes siguiendo los criterios de clasificación descritos en la sección 3.3. Un subconjunto de las asignaciones más representativas se detalla a continuación.

- CR2: "*La información se produce utilizando juicios subjetivos...*"
  - PE1 – Ingreso por equivalencias: una persona puede solicitar el ingreso por equivalencias a una carrera universitaria, esto es posible sólo si la persona interesada cuenta con asignaturas aprobadas en otras carreras de grado. El otorgamiento de la aprobación de cada asignatura en la carrera destino se basa en la evaluación de la documentación presentada y la opinión de los docentes asignados para realizar la evaluación. No existen lineamientos claramente definidos ni el soporte adecuado para que el docente pueda evaluar de manera adecuada.
  - PE2 – Aplicación de normativas: durante las reuniones del Consejo Departamental surgen situaciones que deben dirimirse de acuerdo con el criterio de los consejeros basándose en la normativa vigente. Las ordenanzas y resoluciones se encuentran digitalizadas y son accesibles mediante su identificador, pero no existe un soporte brinde acceso inteligente a las mismas, por ej. cuándo se debe aplicar cierta normativa.
- CR4: "*Los grandes volúmenes de información almacenada dificultan el acceso...*"
  - No se detectaron problemas relevantes vinculados con el volumen de los datos.
- CR7: "*El análisis de contenido automatizado en todas las colecciones...*"
  - PE2 – Descrito anteriormente.

This work is licensed under a Creative Commons "Attribution-NonCommercial-NoDerivatives 4.0 International" license.

Gaona, G., Bollati, V. A., Caballero Muñoz-Reja, I., Gualo, F.: Diseño de un sistema de gobierno de datos: un marco de trabajo para la priorización de procesos. In: Durán Toro, A. (ed.) Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2023). Sistedes (2023). https://hdl.handle.net/11705/JISBD/2023/6178

- PE3 – Seguimiento en línea del desempeño de los alumnos basándose en el análisis de tendencias a través del campus virtual.

La matriz de vinculación entre problemas específicos y causas raíz asociadas resultantes se muestra en la Tabla 5.

Tabla 5. Matriz de vinculación aplicada

	CR2	CR4	CR7
PE1	X		
PE2	X		X
PE3			X
Cantidad	2	0	2

4.2 Clasificación de los procesos de MAMD y las causas raíz probables

La identificación de qué procesos de MAMD deben implantarse como base para resolver, paliar o evitar las causas raíz fue encargada a tres expertos, esto se realizó de acuerdo con lo especificado en la sección 3.4.

Un subconjunto de los procesos de MAMD fueron elegidos para ser clasificados en este ejemplo. Estos fueron consensuados con los expertos con el objetivo de mantener un número manejable de procesos por las restricciones de espacio.

En consecuencia, para este ejemplo se incluyeron todos los procesos del nivel de madurez 1 y 2, estos son: *"Procesamiento del dato"* (ProcDat), *"Gestión de la infraestructura tecnológica del dato"* (InfraTec), *"Gestión de requisitos del dato"* (ReqDat), *"Gestión de configuración del dato"* (ConfDat), *"Gestión del dato histórico"* (DatHis), *"Gestión de seguridad del dato"* (SegDat), *"Gestión del metadato"* (MetDat), *"Control y monitorización de calidad del dato"* (CtrlDQ) y *"Establecimiento de políticas, buenas prácticas y procedimientos del dato"* (PolDat).

Cada experto se encargó de generar la matriz de clasificación de manera separada, y luego se integraron los resultados. La matriz resultante se muestra en la Tabla 6.

Tabla 6. Matriz de clasificación aplicada

	CR2	CR4	CR7
ProcDat	O; M	O; M	O; M
InfraTec	S; S	S; M	S; S
ReqDat	S; C	S; C	O; M
ConfDat	S; W	S; C	S; C
DatHis	S; C	O; S	S; C
SegDat	S; W	O; C	O; C
MetDat	S; W	S; C	O; M
CtrlDQ	E; S	E; S	E; S
PolDat	E; S	E; S	E; M

4.3 Obtención de los procesos priorizados

La matriz de priorización permitirá calcular la prioridad de cada proceso de MAMD respecto a su aportación a la solución de cada causa raíz. Tal como se describió en la sección 3.4, esta matriz se construye sobre la matriz de clasificación (Tabla 6), que es

This work is licensed under a Creative Commons "Attribution-NonCommercial-NoDerivatives 4.0 International" license.

Gaona, G., Bollati, V. A., Caballero Muñoz-Reja, I., Gualo, F.: Diseño de un sistema de gobierno de datos: un marco de trabajo para la priorización de procesos. In: Durán Toro, A. (ed.) Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2023). Sistedes (2023). https://hdl.handle.net/11705/JISBD/2023/6178

genérica, y se le agregan la cantidad de problemas particulares asignados a cada causa raíz (Tabla 5), más la columna de priorización (PP).

La Tabla 7 muestra la matriz de priorización terminada, dónde se aprecia cada prioridad calculada. Cómo ejemplo se toma el proceso DM01, si se aplica la ecuación 1 y se reemplazan los valores correspondientes se obtiene 36, el cálculo se repite para el resto de los procesos involucrados.

Tabla 7. Matriz de priorización aplicada

	CR2	CR4	CR7	PP
Cantidad	2	0	2	
ProcDat	O; M	O; M	O; M	36
InfraTec	S; S	S; M	S; S	16
ReqDat	S; C	S; C	O; M	22
ConfDat	S; W	S; C	S; C	6
DatHis	S; C	O; S	S; C	8
SegDat	S; W	O; C	O; C	8
MetDat	S; W	S; C	O; M	20
CtrlDQ	E; S	E; S	E; S	8
PolDat	E; S	E; S	E; M	10

Los resultados obtenidos con este ejemplo de aplicación de MPPM han sido valorados positivamente por los expertos que participaron de este proceso.

Se pudo constatar que la prioridad calculada para cada proceso es coherente con los niveles de madurez propuestos por MAMD. Sin embargo, la prioridad obtenida para MetDat y ReqDat permite poner el foco en procesos que podrían ser de utilidad para la resolución de los problemas específicos encontrados en el dominio bajo estudio.

Por último, a pesar de los resultados alentadores, la aplicación de MPPM en la UTN-FRRe fue acotada y presenta limitaciones, por lo que estos resultados no pueden ser generalizados a otros dominios, y se pretende extender la comprobación de MPPM mediante el diseño de un caso de estudio más amplio.

5 Conclusiones y trabajos futuros

Los problemas de calidad de datos suelen requerir un abordaje a nivel organizacional mediante el gobierno de datos. Específicamente, la implantación de un SGD y, por ende, la de sus procesos, debe hacerse de manera planificada y progresiva, considerando las particularidades de cada organización.

En este sentido, el Marco para la Priorización de Procesos de MAMD (MPPM) descrito en el presente artículo, permite identificar aquellos procesos de MAMD más importantes que ayudarían a resolver, paliar o evitar las causas raíz y sus problemas asociados que son particulares a un dominio.

El dominio usado para la prueba de concepto de MPPM, ha sido la UTN-FRRe. Esto ha servido como una primera aproximación para comprobar el desempeño de MPPM y ha permitido detectar puntos de mejora potenciales.

Como trabajos futuros se pretende ampliar MPPM analizando la posibilidad de incluir nuevas causas raíz, ajustando los criterios de asignación y las variables

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.

Gaona, G., Bollati, V. A., Caballero Muñoz-Reja, I., Gualo, F.: Diseño de un sistema de gobierno de datos: un marco de trabajo para la priorización de procesos. In: Durán Toro, A. (ed.) Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2023). Sistedes (2023). https://hdl.handle.net/11705/JISBD/2023/6178

contempladas en la priorización, y finalmente con casos de estudios reales validar los resultados obtenidos mediante la implantación de un SGD aplicando MPPM.

Agradecimientos

Esta investigación ha sido parcialmente financiada por los siguientes proyectos y organizaciones: AETHER-UCLM: Una aproximación holística de Smart data para el análisis de datos guiado por el contexto centrada en la calidad y la seguridad (Ministerio de Ciencia e Innovación, PID2020-112540RB-C42), Gobierno del Dato en AAPP del NEA: Análisis, Adaptación e Implementación de Modelos de Gobierno y Calidad de Datos en las Administraciones Públicas de la Región NEA (Universidad Tecnológica Nacional, SITCRE0008590) y por el CONICET. Y ADAGIO (Alarcos' DAta Governance framework and systems generatIOn) (SBPLY/21/180501/000061), financiado por la Consejería de Educación, Cultura y Deportes de la Junta de Comunidades de Castilla-La Mancha)

Referencias

1. van den Broek, T., van Veenstra, A.F.: Governance of big data collaborations: How to balance regulatory compliance and disruptive innovation. Technol. Forecast. Soc. Change. 129, 330–338 (2018). https://doi.org/10.1016/j.techfore.2017.09.040.
2. ISO: ISO 37000:2021 - Governance of organizations. Guidance. (2021).
3. Zhang, Q., Sun, X., Zhang, M.: Data Matters: A Strategic Action Framework for Data Governance. Inf. Manag. 59, 103642 (2022). https://doi.org/10.1016/j.im.2022.103642.
4. Benfeldt Nielsen, O.: A Comprehensive Review of Data Governance Literature. Issue Nr. 8, 3 (2017).
5. UNE: Especificación UNE 0077:2023, Gobierno del Dato. , Madrid (2023).
6. Caballero, I., Piattini, M., Gualo, F.: Marco metodológico para la creación, implantación y mantenimiento de Sistemas de Gobierno de Datos. In: Goñi Sarriguren, A. (ed.) JISBD2022. SISTEDES (2022).
7. Carretero, A.G., Gualo, F., Caballero, I., Piattini, M.: MAMD 2.0: Environment for data quality processes implantation based on ISO 8000-6X and ISO/IEC 33000. Comput. Stand. Interfaces. 54, 139–151 (2017). https://doi.org/10.1016/j.csi.2016.11.008.
8. Weber, K., Otto, B., Österle, H.: One Size Does Not Fit All---A Contingency Approach to Data Governance. 1, 1–27 (2009). https://doi.org/10.1145/1515693.1515696.
9. Strong, D.M., Lee, Y.W., Wang, R.Y.: 10 potholes in the road to information quality. Computer (Long. Beach. Calif). 30, 38–46 (1997). https://doi.org/10.1109/2.607057.
10. Chidichimo, G., Niz, J., Sánchez Cestona, J., Tovillas, P.: Evaluación externa de instituciones universitarias. Estudio sobre las recomendaciones para el mejoramiento institucional. CONEAU-Comisión Nacional de Evaluación y Acreditación Universitaria, Ciudad Autónoma de Buenos Aires (2022).
11. Liliana, L.: A new model of Ishikawa diagram for quality assessment. In: IOP Conference Series: Materials Science and Engineering. Institute of Physics Publishing (2016). https://doi.org/10.1088/1757-899X/161/1/012099.
12. DAMA International: Data Management Body Of Knowledge Book 2. Technics Publications; Second edition, Basking Ridge, New Jersey (2017).

This work is licensed under a Creative Commons "Attribution-NonCommercial-NoDerivatives 4.0 International" license.

Gaona, G., Bollati, V. A., Caballero Muñoz-Reja, I., Gualo, F.: Diseño de un sistema de gobierno de datos: un marco de trabajo para la priorización de procesos. In: Durán Toro, A. (ed.) Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2023). Sistedes (2023). https://hdl.handle.net/11705/JISBD/2023/6178

13. Soares, S.: The IBM data governance unified process: driving business value with IBM software and best practices. MC Press Online (2010).
14. EDM Council: Data Management Capability Assessment Model - DCAM, https://dgpo.org/wp-content/uploads/2016/06/EDMC_DCAM_-_WORKING_DRAFT_VERSION_0.7.pdf, last accessed 2023/05/06.
15. Alarcos Research Group: Modelo Alarcos de Mejora de Datos v.3.0, http://mamd.dqteam.es, last accessed 2022/05/06.
16. UNE: Especificación UNE 0078:2023, Gestión del Dato. , Madrid (2023).
17. UNE: Especificación UNE 0079:2023, Gestión de la Calidad del Dato. , Madrid (2023).
18. Melão, N., Pidd, M.: A conceptual framework for understanding business processes and business process modelling. Inf. Syst. J. 10, 105–129 (2000). https://doi.org/10.1046/j.1365-2575.2000.00075.x.
19. Hatton, S.: Early prioritisation of goals. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 235–244 (2007). https://doi.org/10.1007/978-3-540-76292-8_29.
20. Maldonado, M., Mugglestone, K., Roberson, A.J.: Developing a Data-Informed Campus Culture: Opportunities and Guidance for Institutional Data Use. , Washington (2021).

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.