

**Universidad Central “Marta Abreu” de Las Villas
Facultad de Matemática, Física y Computación**



Maestría en Ciencia de la Computación

***Procedimiento para la Gestión de la Calidad de Datos
del Sistema Informativo Bancario***

Tesis presentada en opción al Título Académico de Máster en Ciencia de Computación

Autor: Lic. Verónica Evora Torres

Tutor: Dr. C. Ramiro A. Pérez Vázquez

2012

*“Los que se enamoran de la práctica sin la teoría
son como los pilotos sin timón ni brújula,
que nunca podrán saber a dónde van”.*

*“He estado impresionado con la urgencia de hacer.
Conocimiento no es suficiente, debemos aplicarlo.
Estar deseoso no es suficiente, tenemos que crear”.*

- Leonardo Da Vinci

Dedico este trabajo a las personas que siempre me han dado la confianza que he necesitado.

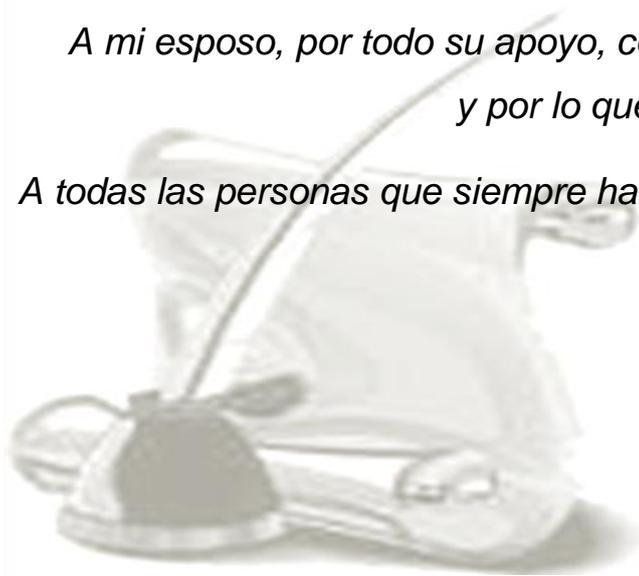
A mi familia por su hermosa compañía y estar siempre pendientes de mi.

A mi madre, por estar siempre, por todos los sacrificios que has hecho para que yo este hoy aquí, por dedicarme tu vida entera.

A mi hijo, que fue mi inspiración.

A mi esposo, por todo su apoyo, comprensión, ayuda y por lo que tuvo que esperar.

A todas las personas que siempre han estado a mi lado.



Agradezco a las personas que han contribuido al desarrollo de este trabajo.

A mi tutor, el Dr. Ramiro Pérez Vázquez por el apoyo que me ha brindado, por su ayuda y empeño en este trabajo, el cuál sin su esfuerzo no se hubiera logrado.

A Lourdes por sus ideas y a Juan Carlos, por todo lo que me dieron a pesar de la distancia y estar siempre conmigo en los momentos difíciles.

A Linnett, Yadiel y Sonia, mis compañeros de trabajo, por apoyarme.

El **Banco Central de Cuba** (BCC) como organismo rector, tiene la misión de velar por la estabilidad de la banca, supervisar el resto del sistema bancario, proponer e implementar un sistema de pagos y una política monetaria conveniente para un crecimiento económico sostenido. Para poder asumir todas estas funciones el BCC necesitó de un Sistema Informativo Bancario (SIB) abierto a la incorporación de nuevas funciones propias de su actividad y respaldado legalmente, capaz de asimilar el cúmulo de datos procedentes de los organismos de la Administración Central del Estado y de las instituciones del Sistema Bancario Financiero Nacional.

Los datos **son un activo crítico** para cualquier organización, pero si carecen de calidad, están incompletos o son poco eficientes, aunque se tenga el mejor y más caro sistema informativo, la propia información se proporcionará con calidad pésima, por tanto, poseer una visión global de los datos y su entorno nos garantiza el análisis y valoración de los mismos en el actual SIB.

El objetivo general del trabajo es elaborar un procedimiento para efectuar la evaluación y gestión de los datos financieros en el SIB, mediante un apoyo metodológico que defina la visión, los requisitos de la organización y la planificación necesaria para trazar un conjunto coherente de políticas, “buenas prácticas” y acciones de mejoras que contribuyan a elevar la calidad de los mismos. Para lograr el objetivo, se realizó el estudio del estado del arte respecto a metodologías de calidad de datos y se validó el procedimiento propuesto en el objeto de estudio, como vía de comprobación y factibilidad de la investigación.

The Central Bank of Cuba as a ruling main office has the mission to run the stability of the banking system, supervise the rest of it, propose and introduce a payment system and a convenient monetary policy for a sustained economic growth. Being able to assume all these functions, The Central Bank of Cuba was in need to create a Banking Informative System (SIB) opened to the introduction of its own new activity functions and back legally up, able to control an amount of data coming from The Central State-controlled Institutions and The National Financial System as well.

The data are a critical assets for any organization, but if they scarce of quality, they're incomplete or they're less efficient although the best and the most expensive informative system is available, the information itself will be given in a very bad quality, therefore, having a global vision of the data and its environment guarantees the analysis and valuation of it in the current SIB.

The general goal of the work is creating a procedure to make the evaluation and management of the financial data in the SIB, with a methodological support that defines the vision, the requirement of the organization and the necessary planning to draw up a group of coherent policies, "good practical" and actions of improvements that contribute to increase the quality of them. For fulfilling this goal, the state of art study concerning quality data methodologies was made and the procedure suggested in the object of study was validated, as a way of verification and feasibility of the investigation.

INTRODUCCION	1
CAPÍTULO I. MARCO TEÓRICO- REFERENCIAL	8
1.1 <i>Ciclo de vida de la información</i>	11
1.2 <i>Dimensiones de calidad de datos</i>	12
1.3 <i>Técnicas de impacto</i>	15
1.4 <i>Categorías y especificaciones de los datos</i>	16
1.5 <i>Metodologías y Marcos de trabajos de calidad de datos</i>	17
1.5.1 <i>TIQM: Gestión total de la calidad de la información</i>	19
1.5.2 <i>Conocimientos de Dirección de Empresa: El enfoque de Calidad de Datos</i>	21
1.5.3 <i>Metodología para la valoración de Calidad de Datos en los datos Financieros</i>	23
1.5.4 <i>CDQM: Calidad de Datos - Conceptos, Metodologías y Técnicas</i>	24
1.5.5 <i>Diez pasos hacia una Calidad de Datos y de confianza de Información</i>	26
1.5.6 <i>Caracterización de Marcos de Trabajo y Metodologías</i>	28
1.5.7 <i>Herramientas y productos líderes en la Calidad e Integración de los datos</i>	30
1.6 <i>Conclusiones parciales</i>	32
CAPÍTULO II. PROCEDIMIENTO PARA LA GESTIÓN DE LA CALIDAD DE DATOS EN EL SIB	34
2.1 <i>Evaluación del modelo general del actual sistema SIB</i>	34
2.2 <i>Mejores prácticas y pautas para la estructuración del procedimiento</i>	36
2.3 <i>Descripción de los pasos o etapas del procedimiento</i>	37
2.3.1 <i>Etapa 1. Definir necesidades y especificaciones</i>	38
2.3.2 <i>Etapa 2. Analizar entorno de la información</i>	43
2.3.3 <i>Etapa 3. Evaluar calidad de los datos</i>	48
2.3.4 <i>Etapa 4. Evaluar impacto</i>	54
2.3.5 <i>Etapa 5. Identificar causas raíz</i>	56
2.3.6 <i>Etapa 6. Plan de mejoras</i>	57
2.3.7 <i>Etapa 7. Plan de comunicación</i>	58
2.4 <i>Conclusiones parciales</i>	59
CAPÍTULO III. EVALUACIÓN DEL PROCEDIMIENTO EN EL ACTUAL SIB	61
3.1 <i>Evaluación del Procedimiento en el SIB</i>	61
3.1.1 <i>Etapa 1. Definir necesidades y especificaciones</i>	61
3.1.2 <i>Etapa 2. Analizar entorno de la información</i>	63
3.1.3 <i>Etapa 3. Evaluar calidad de los datos</i>	67
3.1.4 <i>Etapa 4. Evaluar impacto</i>	73
3.1.5 <i>Etapa 5. Identificar causas raíz</i>	76
3.1.6 <i>Etapa 6. Plan de mejoras</i>	78
3.1.7 <i>Etapa 7. Plan de comunicación</i>	78

3.2	<i>Aplicabilidad de las “herramientas” y productos propuestos</i>	79
3.3	<i>Conclusiones parciales</i>	80
CONCLUSIONES Y RECOMENDACIONES		81
BIBLIOGRAFIA		84
ANEXOS		92

INTRODUCCION

El Sistema Informativo del Banco Central de Cuba (SIB) es un sistema operacional diseñado para la captación, almacenamiento y distribución centralizada de información (modelos informativos elaborados según metodologías establecidas al efecto), la cual se dirige al Banco Central desde los Organismos de la Administración Central del Estado (OACE) y las instituciones del Sistema Bancario Financiero Nacional (SBFN) y es procesada en un servidor que funciona las 24 horas del día. El SIB está respaldado legalmente por el decreto-Ley 172 (artículos 29 y 36) y la resolución del 99/2001 del BCC (Guerrero, Julio, 2006).

Durante años desde la creación del Sistema Informativo en el BCC, la Dirección de Estadísticas Monetarias y Financieras (DEMF) ha batallado con los problemas de calidad de datos. Datos financieros deficientes fueron obstaculizando la entrega de informaciones oportunas y confiables a diferentes instituciones del Estado, empleándose demasiado tiempo y esfuerzo en corregirlos. A medida que el proyecto se expandía, la DEMF¹ centralizó la administración de datos y elaboró metodologías para la confección y entrega de los modelos informativos que constituyen la entrada de información al SIB y están diseñados en función de las necesidades informativas. Existen actualmente más de 42 modelos (Anexo 1) de los cuales 30 deben ser reportados por los bancos, 14 por las instituciones financieras no bancarias y 10 por los OACE, con fechas de cumplimientos dispersas en el transcurso del mes evitando concentración en el tiempo; siendo 24 modelos de uso compartido entre las direcciones del BCC.

Datos estandarizados y optimizados producen una significativa contribución al valor añadido de los objetivos estratégicos y operativos de la organización. Asimismo, si se cuenta con un procedimiento para la mejora de la calidad de la información, se conseguirá el mayor beneficio posible a los analistas para el control de situaciones que deben ser notificadas y atendidas.

¿Qué pasos necesitaría emprender la alta dirección del banco para resolver los problemas de calidad de datos acorde a sus necesidades?

¹ Tiene como misión, dotar al BCC de la información estadística necesaria para el cumplimiento de sus funciones evalúa el desempeño del Sector Externo y la evolución del Financiamiento del SBFN a la economía interna.

- Comprender los costos legítimos y las causas de la mala calidad de los datos.
- Establecer una estrategia definida para valorar la calidad de los mismos.
- Alinear expectativas de la organización y del departamento de Informática.
- Decidir la metodología adecuada en función del nivel de apoyo logrado por la alta dirección.
- Asegurar que los objetivos impulsan la funcionalidad.
- Comprender los gastos de desarrollar la solución propietaria.
- Disponer de personal para la evaluación y análisis de la calidad de datos.
- Valorar Costo – Beneficio.

Existen varias definiciones en la literatura asociada al concepto de Calidad de Datos:

- La Calidad de Datos no sólo se refiere a la ausencia de defectos, los datos deben proporcionar una visión única,
- Es el estado de integridad, la validez, la consistencia, la frescura y la exactitud que hacen el dato apropiado para un uso específico,
- La totalidad de rasgos y características de los datos que llevan a satisfacer un propósito dado.

El concepto Calidad de la Información está surgiendo con fuerza en los últimos años, se trata del concepto de Calidad de Datos orientado a los Sistemas de Información, es decir, a la mejora de la información proporcionada a la compañía u organización (Porras, Mínguez A., 2008). Algunos expertos extienden el término a lo que también se denomina “Calidad de Metadatos” y surge en grandes corporaciones que cuentan con miles de atributos e indicadores. En lo adelante, esta es la definición que utilizaremos en la investigación, propuesta por Danette McGilvray en entrevista (citada en Dylan, Jones, 2009d) y ampliada por otros autores:

“La Calidad de la Información es el grado en que la información y los datos pueden ser una fuente confiable para cualquiera que requiera su uso; de forma completa, correcta, consistente, oportuna para las personas que toman decisiones, accesible a los clientes y dirigida a la organización para conseguir sus objetivos”, por tanto se considera buena si los datos (Olson, Jack E., 2003; Montero, L., 2004; Porras, Mínguez A., 2008) son:

Precisos: sin errores introducidos en la entrada de datos, en su transformación, almacenamiento y distribución.

Completos: disponible para todos en las bases de datos a que se refiere y/o para aquellos donde son necesarios.

Coherentes: en su definición y tratamiento, a través de los sistemas de información y bases de datos de la organización.

Subordinados: a las reglas de negocio de la organización.

Enriquecidos: incrementados con datos interesantes de fuentes internas o externas que mejoran el valor general de los datos.

Mantenidos y documentados: para conservar su calidad a través del tiempo, permiten realizar análisis apropiados y ser reutilizables en otros sistemas o procesos.

Contar con datos de calidad, entre otras cosas, permite: suministrar datos precisos y completos que conllevan a mejores resultados y a un alto rendimiento aumentando la confianza del usuario, al garantizar que pase más tiempo analizando los datos y menos tiempo comprobando la calidad de las fuentes de datos; aumenta el conocimiento de los datos para conocer dónde se están produciendo los errores y corregirlos de manera proactiva; construir mejores sistemas de información pues las iniciativas de calidad de datos identifican las áreas problemáticas, que unidas a la gestión continua de la calidad soportan los ajustes y cambios de requisitos funcionales para ampliar la vida útil del sistema sin necesidad de mantener sistemas operacionales paralelos e independientes al disponer de una fuente de datos limpia y consistente.

Entre las causas de la mala calidad de datos de las fuentes externas del SIB se pueden señalar:

Falta de motivación: Las personas responsables de la entrega de datos, debido a la falta de incentivos individuales a la producción de datos con calidad, originan información dispersa e inconsistente.

Cambio organizativo: Los nuevos requerimientos de datos, hacen que contingencias ocultas se manifiesten en la forma de problemas más serios (Ejemplo: la principal causa de problemas es la “introducción manual de los datos”); algunos comprenden o reconocen que los datos son un activo estratégico importante, pero la mayoría no emplea procedimiento alguno de control de calidad, ni hace uso de las herramientas necesarias para garantizar la coherencia y precisión de los datos.

Falta de normas de calidad y controles: Hay carencia de un procedimiento de calidad de datos, ausencia de normas de calidad de datos o no conformidad si las normas existen, coexiste una mala conversión de datos y tratamiento inconsistente de los mismos desde otros sistemas fuentes, así como falta de documentación y evaluación periódica.

En muchas ocasiones se delega la responsabilidad de la calidad de los datos mayoritariamente en el departamento de Informática; área ésta que aunque implicada, no tiene la capacidad de modificar los procesos del negocio o comportamientos fuera de su área de influencia que permitan la mejora sustancial de la calidad de los mismos. Un principio fundamental en la gestión de Calidad de Datos es la detección y corrección de los errores lo más cerca posible de su origen, con el objetivo de minimizar los costos asociados.

En la actualidad, en América Latina, el acceso a la literatura se obtiene a través de Internet. Sin embargo, no existen programas de capacitación, la presencia de instituciones y asociaciones internacionales es escasa y el problema más importante es que no hay cultura de calidad de los datos (Dylan, Jones, 2012). Esto está creciendo de forma reactiva a través de los sectores controlados tales como los financieros. La calidad de los datos es vista como una opción, se toman acciones con visión a corto plazo. En los sectores menos controlados por la necesidad de gestionar la iniciativa de calidad de datos proviene de las habilidades de liderazgo ejecutivo, que generalmente pertenecen al área de marketing y en algunos casos en las áreas de informática.

En nuestro país, la literatura refleja que se realizan diferentes proyectos en este campo. Un elemento importante considerado en estos proyectos es el diagnóstico, en ese sentido también se han desarrollado procedimientos o metodologías para el diagnóstico de la calidad de los datos que, entre otros aspectos, inciden en la búsqueda de las causas que provocan los problemas de calidad (Pérez, D., 2008), (Vilalta, 2008), (Heredia, J. y Vilalta J., 2009), (Pérez, D y Vilalta, J., 2010b).

En la construcción del marco teórico de la investigación se observa que las metodologías y/o procedimientos relacionados con la calidad de datos, en general son abstractos y de alto nivel, no contienen la suficiente narrativa, no son completamente funcionales o les falta integración con el negocio; lo que unido a la **situación problemática** que se deriva de la ausencia de un procedimiento con fundamento metodológico en el Sistema Informativo

Bancario para la gestión de la calidad de los datos financieros, justifican el planteamiento del **problema científico**:

Se hace necesario elaborar un procedimiento estructurado y formalizado en el Sistema Bancario que nos potencie la selección de mejores prácticas reconocidas para la gestión de la calidad de los datos.

El propósito primordial de este trabajo tiene como **objetivo general**:

Elaborar un procedimiento metodológico para la gestión de la calidad de los datos con un enfoque proactivo y estratégico;

*La consecución de este objetivo principal se lleva a cabo a través de otros **objetivos específicos** que se pueden considerar como subordinados a este. Se proponen los siguientes objetivos concretos:*

- ✓ Construir el marco teórico-referencial de la investigación derivado de la consulta actualizada de la literatura sobre la calidad de datos, enfatizando en las características de las metodologías y marcos de trabajo, de modo que admitan la adaptación a la solución del SIB.
- ✓ Diseñar un procedimiento coherente con los objetivos del negocio, que permita obtener y formalizar las especificaciones de calidad de los datos en el SIB de manera completa y formal, así como evaluar la aplicabilidad de herramientas y técnicas para lograr la calidad deseada.
- ✓ Verificar la funcionalidad del procedimiento propuesto de modo que permita identificar las causas raíces de los problemas de calidad y definir el mejoramiento de los procesos en el SIB para elevar la satisfacción de los clientes.

De acuerdo a lo anteriormente planteado, así como de la revisión de la bibliografía especializada y otras fuentes en la construcción del marco teórico o referencial de la investigación, se formuló la **hipótesis** siguiente:

Con la concepción e implementación de un procedimiento metodológico se revelará dónde se están produciendo los errores en el SIB, aumentará el conocimiento de los datos y se mejorará la gestión de la calidad de los mismos, así como, la posibilidad de medir los beneficios del procedimiento desarrollado.

La **novedad científica** estará dada por la elaboración de un procedimiento metodológico que permita definir la calidad de los datos en el SIB, agregando valor a la organización con la mejora de los procesos de calidad de datos y las herramientas apropiadas mediante una visión conceptual única y completa.

El presente trabajo tiene valor teórico ya que garantiza la creación del procedimiento bajo una perspectiva holística, incrementando el conocimiento de los datos para constituir un plan de acción que ayude propiamente a una correcta disciplina informativa con un apoyo metodológico; a su vez tiene valor práctico, mediante la descripción y detalle del procedimiento en sí, que favorece la mejora de la calidad de la información mediante la integración de las mejores prácticas, el cual será detallado en el desarrollo del trabajo. Por otro lado, es conveniente advertir y/o corregir datos inconsistentes, incorrectos, incompletos o ambiguos evitando el posible manejo de datos no confiables; evitar el desarrollo de sistemas computacionales paralelos e independientes sin coordinación con la DEMF, o las peticiones de informaciones que luego se hacen permanentes sin un soporte metodológico descrito; y finalmente, para conseguir medir y compulsar la efectividad de los tributantes en el cumplimiento de los compromisos informativos con el BCC.

Para explicar el desarrollo de cada uno de los objetivos se estructuró este documento en tres capítulos.

En el capítulo 1, "*Marco Teórico – Referencial*" se exponen los resultados más relevantes que recoge la literatura relacionados con metodologías y marcos de trabajo, herramientas, técnicas y conceptos para la calidad de datos, tomando partido en los aspectos que resultan bondadosos para la elaboración del procedimiento que se propone.

En el capítulo 2, "*Procedimiento para la gestión de la calidad de datos en el SIB*", se ofrece una amplia descripción de la definición del procedimiento elaborado sobre la adaptación de las metodologías y marcos de trabajo valorados, haciendo énfasis en cada una de las etapas o pasos.

En el capítulo 3, "*Evaluación del procedimiento en el actual SIB*", se ofrece la aplicación y validación del procedimiento elaborado.

Finalmente completan la tesis un cuerpo de Conclusiones y Recomendaciones derivadas de la investigación realizada, la Bibliografía consultada y un grupo de Anexos como

complemento de los aspectos expuestos.

CAPÍTULO I. MARCO TEÓRICO- REFERENCIAL

En este capítulo se aborda la revisión de la literatura especializada, así como otras fuentes bibliográficas y referenciales consultadas, se estructuró de forma tal que permitiera el análisis general de un grupo de metodologías y *frameworks*² disponibles para la calidad de datos, la terminología y los conceptos claves que son importantes para comprender e interpretar la particular situación de la calidad de datos, permitiendo sentar las bases teóricas y prácticas de la investigación.

Los datos son la materia prima de la información. El conocimiento obtenido de los mismos proporciona los requisitos básicos para una mejor definición de la estrategia y operaciones de negocio (Schober, 2009), (Uniserv, 2005). La Guía DAMA-DMBOK (*Data Management Book of Knowledge*) presenta una visión estándar de la industria de las funciones de gestión de datos, terminología y mejores prácticas, sin detallar los métodos y técnicas específicas (Mark, Mosley, 2008). Ayuda a crear un consenso dentro de la comunidad de gestión de datos y a entender las responsabilidades de los administradores y los profesionales de datos. Proporciona la base para la evaluación de la eficacia de la gestión de datos y la madurez, mediante la orientación de los esfuerzos para aplicar y mejorar las funciones de gestión de datos. Señala a los lectores otras fuentes de conocimiento y sugiere áreas de investigación en el campo de la gestión de datos ayudando a las organizaciones en su estrategia. Los modelos de referencia para la Guía DAMA-DMBOK son la Dirección de Proyectos del Conocimiento (PMBOK).

- Datos de Administración y Gobierno ha pasado a denominarse simplemente **Data Governance**, en el reconocimiento de actividades de gestión de datos y las responsabilidades.
- Mejoramiento de la Calidad de Datos se ha cambiado el nombre a **Gestión de Calidad de Datos**, que refleja el alcance más amplio de las actividades dentro de la función.

² En el presente trabajo utilizaremos indistintamente la palabra “*framework*” de origen inglés, como marco de trabajo

- La dirección de Calidad de Datos ha sido cambiada a **Reference & Master Data Management**, reconociendo esta como una forma especializada de la dirección de calidad.
- Gestión de metadatos ha sido renombrado a **Meta Data Management** con respecto a restricciones de denominación.

La figura 1.1 muestra las diez funciones de la gestión de datos:

Gobierno de los datos: El ejercicio de la autoridad, el control y la toma de decisiones compartida (planificación, control y aplicación) a través de la gestión de los activos de datos. El gobierno es de alto nivel de planificación y control de la gestión de datos.

Arquitectura de administración de datos: El desarrollo y mantenimiento de la arquitectura de datos, en el contexto de toda la arquitectura de la empresa y su relación con las soluciones de aplicación del sistema y proyectos que implementan arquitectura empresarial.

Datos para el Desarrollo: Los datos son enfocados en las actividades de desarrollo del sistema y su ciclo de vida, incluyendo el modelado y el análisis de requisitos, el diseño, implementación y mantenimiento de bases de datos para la solución.

Gestión de base de datos de Operaciones: Planificación, control y apoyo a la estructura de datos en todo el ciclo de vida de los mismos, desde la creación y la adquisición hasta el archivo.

Gestión de Seguridad de datos: Las actividades de planificación, ejecución y control que garantizan la privacidad, la confidencialidad y evita que personal no autorizado e inadecuado tenga acceso a los datos para la creación o cambio.

Referencia y Gestión de Datos Maestros: La planificación, ejecución y control de actividades para garantizar la coherencia de los valores de los datos contextuales.

Almacenamiento de Datos y Gestión de Inteligencia del Negocio: Planificación, procesos de implementación y de control para proporcionar datos de apoyo a las decisiones en informes, consultas y análisis.

Gestión de Documentos y Contenidos: Planificación, ejecución y control de actividades para almacenar, proteger y acceder a los datos que se encuentran en archivos electrónicos y registros físicos (incluyendo textos, gráficos, imágenes).

Gestión de Metadatos: La integración, el control y la entrega de los metadatos.

Gestión de Calidad de datos: Las actividades de planificación, ejecución y control que aplican las técnicas de gestión de calidad para medir, evaluar, mejorar y asegurar la conveniencia de los datos para su uso.

Cada una de estas funciones se descompone a su vez en actividades.

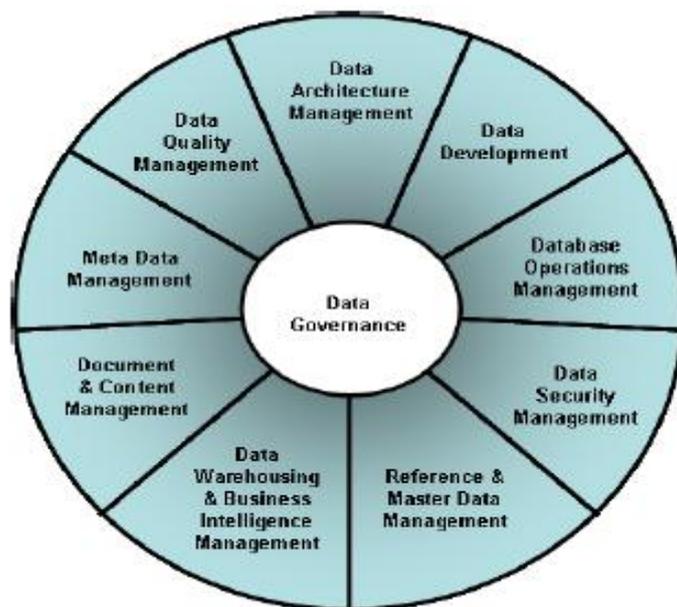


Figura 1.1 Funciones de la gestión de datos. Tomado de © 2008 DAMA Internacional – Derecho Reservado

Existen diferentes enfoques para medir la calidad de los datos (Redman, Thomas C., 2004). Algunos autores se centran en la definición de dimensiones de calidad en diferentes contextos de aplicación (Wang, R.& Strong, D., 1996), (Batini, Carlo & Scannapieca, Monica, 2006), (Pipino, L., Yang W. Lee, and Wang Richard, 2002). Otros autores estudian cómo medirla en forma práctica, por ejemplo, en (Etcheverry, L., Tercia, S., Marotta, A. y Peralta, V., 2007) se mide la exactitud de datos en un sistema de Almacén de Datos. En un contexto de Sistemas de Integración de Datos (Wikipedia, 2009), aparece también la necesidad de evaluar la calidad de datos resultante de sofisticados procesos de integración

que combinan fuentes de datos de calidad muy variada. Algunos autores se centran en la medición de la calidad de los datos que responden a consultas de usuarios y la satisfacción de las exigencias de dichos usuarios en términos de calidad (Peralta, 2006).

Para la concepción e implementación de un procedimiento, es necesario comprender algunos conceptos relacionados con la calidad de datos que serán considerados en los epígrafes sucesivos. Los conceptos fundamentales son: *el ciclo de vida de la información* que consiste en los procesos o fases requeridos para manipular cualquier recurso, *las dimensiones de la calidad de los datos*; aspectos o características de calidad que proporcionan un modo de medir y tratar la calidad de los datos, *las técnicas de impacto en la organización*; medidas cualitativas y cuantitativas para determinar los efectos de la calidad y por último, *las categorías y especificaciones de los datos* para agrupar datos con características comunes e incluir datos estándar, modelos de datos, reglas del negocio o metadatos.

1.1 Ciclo de vida de la información

La información es el valor añadido del producto de datos obtenido por su interpretación (English, Larry, 1999). Para manipular cualquier recurso es importante comprender la idea del ciclo de vida referente a los procesos de cambio y desarrollo a través de la vida útil de algo. La mayor diferencia entre información como recurso y cualquier otro recurso, es su reusabilidad, no es consumida cuando es utilizada.

Tabla 1.1 *Ciclo de vida de la información. Fases y actividades*

Fases del ciclo de vida de la información	Definición	Ejemplo de actividades
Planificar	Preparar el recurso	Identificar objetivos, arquitectura, desarrollar estándares y definiciones para diseñar aplicaciones, bases de datos y procesos.
Obtener	Adquirir el recurso	Cargar datos de ficheros externos, crear registros.
Almacenar	Conservar la información del recurso electrónicamente	Almacenar datos electrónicamente en bases de datos o ficheros.
Mantener	Asegurarse que el	Actualizar, manipular, estandarizar, validar,

	recurso trabaja apropiadamente	transformar o consolidar registros.
Aplicar	Uso del recurso para alcanzar los objetivos	Recuperar datos; usar la información para reportes, toma de decisiones, procesos automáticos.
Disponer	Descartar el recurso cuando no será utilizado	Archivar, eliminar datos o registros.

El ciclo de vida no es un proceso lineal sino iterativo. Existen cuatro importantes componentes que impactan en la calidad de los datos a través de este ciclo.

Datos (Qué) - Conocer hechos u otros aspectos de interés para la organización.

Procesos (Cómo) - Son las funciones, actividades, acciones, tareas o procedimientos que hurtan en los datos.

Organización y Personas (Quiénes) - Equipos, roles, responsabilidades individuales que afectan o usan los datos o están involucrados con los procesos.

Tecnología (Cómo) - Ventanas, aplicaciones, bases de datos, programas, almacenamiento o manipulación de datos involucrados con los procesos.

1.2 Dimensiones de calidad de datos

Son muchos los trabajos que definen, modelan, evalúan o proponen mejoras para la calidad de datos. En general, hay consenso en que la calidad es un concepto multidimensional. La calidad suele estudiarse en múltiples dimensiones que caracterizan diferentes facetas de los datos, por ejemplo, exactitud, accesibilidad, integridad, precisión, confiabilidad, completitud, consistencia, flexibilidad, trazabilidad, seguridad en el acceso, facilidad de manipulación, relevancia, entre otros (Pipino, L., Yang W. Lee, and Wang Richard, 2002), (Wang, R.& Strong, D., 1996), (Motro, A. & Rakov, I., 1998). No obstante, los diferentes trabajos difieren en las dimensiones de calidad que deberían estudiarse. Wang y Strong (1996) afirman que para mejorar la calidad de los datos, es necesario entender qué significa la calidad para los usuarios y presentan una clasificación de dimensiones de calidad que son más relevantes para éstos: relevancia, exactitud, interpretación, accesibilidad, entre otros. Otros autores proponen listas de dimensiones principales para algunos tipos de sistemas o dominios de aplicación, por ejemplo,

completitud, unicidad, consistencia, frescura, exactitud para Sistemas de Integración de datos (Akoka, L., Berti-Équille, O., Boucelma, M. & Buzeghoub, M., 2007), completitud, credibilidad, exactitud, consistencia e interpretación para sistemas de Almacén de datos (Jarke, M., Jeusfeld, M.A., Quix, C. & Vassiliadis, P., 1999), exactitud, completitud, frescura y consistencia para Sistemas Web (Gertz, M., Ozzu, M., Sattler, K. & Saake G., 2004). Otros trabajos se focalizan en el estudio en profundidad de algunas dimensiones, por ejemplo completitud (Naumann, F., Freytag, J. & Leser, U., 2003), frescura (Peralta, V., Ruggia R. & Bouzegoub, M., 2004) o exactitud (Etcheverry, L., Tercia, S., Marotta, A. y Peralta, V., 2007). Tampoco hay consenso en cuanto a definiciones y forma de nombrar las dimensiones de calidad, además, para una misma dimensión podemos encontrar en la literatura diferentes definiciones. Tomemos exactitud como ejemplo, ya que es una de las dimensiones más estudiada. En (Motro, A. & Rakov, I., 1998; Loshin, David, 2001a; Redman, Thomas C., 2001a; Olson, Jack E., 2003) la describen como “la información que está disponible y contiene los valores correctos”; mientras que Wang (1996) da una definición más completa, describiéndola como “los datos son correctos, confiables y certificados como libres de error”.

Cada dimensión de calidad de los datos, requiere diferentes herramientas, técnicas y procesos para medirla (Trillium Software, 2004b), (Matus, C., 2007) y es importante valorar el tiempo y los recursos humanos disponibles para completar su evaluación, por tanto es importante seleccionar aquellas dimensiones que mejor precisamos para nuestra necesidad (McGilvray, Danette, 2008). Diferenciarlas nos ayuda a concordar dimensiones con las necesidades del negocio y priorizar cuáles evaluaciones se completarán primero, comprender que pretendemos hacer para evaluarlas, definir y manejar la secuencia de actividades en el tiempo y las restricciones de los recursos (Olson, Jack E., 2003).

Si bien las dimensiones suelen estudiarse de forma independiente, es ampliamente aceptado que muchas dimensiones están relacionadas y se influyen mutuamente (Martirena, 2008). Algunos trabajos estudian las relaciones entre algunas dimensiones de calidad en sistemas específicos, por ejemplo (Bright, L. & Raschid, L., 2002), (Ballou D. & Pazer H., 1995). En otras palabras, la cantidad de dimensiones de calidad existentes y su interrelación, provoca que el estudio de la calidad de datos sea un problema complejo de varias variables. Se profundizará en algunas dimensiones de calidad que pueden ser relevantes para el objeto de estudio seleccionado.

Tabla 1.2 Dimensiones de calidad de datos

Dimensiones de calidad de datos	Definición (D) y Notas (N)
Especificación de los datos	<p>D - Una medida de la existencia, completamiento, calidad y documentación de datos estándar, modelos de datos, reglas de negocio, metadatos y datos de referencia.</p> <p>N - Proveen el patrón para comparar los resultados de valoración de calidad de datos y la instrucción para ingresar los datos, diseñar programas de carga de datos a mano, poner al día la información y desarrollar aplicaciones.</p>
Integridad de datos Fundamentos	<p>D - Una medida de la existencia, la validez, la estructura, el contenido y las otras características básicas de los datos.</p> <p>N - Las restantes dimensiones se desarrollan sobre lo aprendido con los fundamentos. Incluye medidas como el estado completo, la validez, lista de valores mínimos/máximos, rangos e integridad referencial.</p>
Exactitud	<p>D - Una medida de la corrección del contenido de los datos, exige que una fuente autorizada de referencia sea identificada y accesible.</p> <p>N - La exactitud de datos requiere comparar los datos con el mundo real.</p>
Consistencia y Sincronización	<p>D - Una medida de la equivalencia de la información almacenada.</p> <p>N - La equivalencia es el grado en que datos guardados son conceptualmente iguales.</p>
Disponibilidad y oportunidad	<p>D - Una medida del grado en que los datos están actualizados y disponibles para el uso.</p> <p>N - Los valores de datos cambian en el tiempo y puede existir una brecha hasta que sean actualizados en la base de datos.</p>
Facilidad de uso y conservación	<p>D- Una medida del grado al que los datos pueden ser accedidos y usados y el grado de conservación al que los datos pueden ser actualizados, mantenidos y manipulados.</p>
Cobertura de datos	<p>D- Una medida de la disponibilidad y el alcance de los datos comparados con el universo total de datos o población del interés.</p>
Presentación de datos	<p>D- Mide cómo la información es presentada y coleccionada para quienes la utilizan. Formato y apariencia que respaldan el uso apropiado de la información.</p>
Percepción, relevancia y constancia	<p>D- Una medida de la percepción y confianza en la calidad de los datos: la importancia, valor y relación de los datos para la necesidad de la organización.</p>
Deterioro	<p>D- Una medida del cambio negativo para los datos.</p>
Duplicación	<p>D- Una medida de inesperados datos duplicados a través del sistema para un atributo particular, registro o conjunto de datos.</p>

Las dimensiones de calidad capturan facetas de alto nivel de la calidad de los datos o de los procesos que manipulan esos datos. Un factor representa un aspecto en particular de una dimensión de calidad (Martirena, 2008), por ejemplo, exactitud involucra corrección semántica, corrección sintáctica y precisión (Peralta, 2006). Un factor puede ser más adecuado que otro para algún tipo de problema o aplicación.

1.3 Técnicas de impacto

Las técnicas de impacto en la organización enfocan cómo será empleada la información: completar transacciones, creación de reportes, toma de decisiones, ejecución de procesos automáticos o fuentes de salida para otras aplicaciones (McGilvray, Danette, 2008). Tienen un impacto relacionado con tiempo y complejidad, generalmente a menor tiempo menos complejidad y mayor tiempo más complejas, lo que no significa que las menos complejas sean menos útiles. Las técnicas utilizan medidas cualitativas y cuantitativas (De Amicis, Fabricio & Batini, Carlo, 2004) para determinar los efectos de la calidad.

Tabla 1.3. Técnicas de impacto en la organización

Técnicas de impacto	Definición
Anécdotas	Ponen en orden ejemplos o historias sobre el impacto de la mala calidad.
Tratamiento y uso	Documenta cómo la información es actualmente utilizada y cuáles son los planes futuros.
Matriz de Beneficio contra Costo	Analiza y evalúa la relación entre beneficios y gastos de los problemas, recomendaciones o mejoras.
Clasificación y Prioridad	Clasifica el impacto de los datos faltantes e incorrectos.
Cinco Motivos para El impacto	Preguntar cinco veces el tiempo para alcanzar el impacto real
Costo de la baja calidad	Cuantificar los costos e ingresos para el impacto de la baja calidad
Análisis Costo - Beneficio	Compara los beneficios potenciales de investigar la calidad con la anticipación de los costos. Incluye el retorno de inversión (ROI)

1.4 Categorías y especificaciones de los datos

Las categorías son útiles porque ciertos datos pueden ser tratados basados en su clasificación y características comprendiendo sus relaciones y dependencias.

Tabla 1.4. Categorías de los datos

Categorías de los datos	Definición
Datos Maestros	Describen datos que están involucrados en la organización del negocio y tienden a ser usados por procesos, sistemas, formato de datos estandarizados y valores sincronizados que son críticos para una satisfactoria integración de los sistemas.
Datos Transaccionales	Describen eventos internos o externos que toman lugar en la conducta de la organización. Son típicamente agrupados dentro de registros transaccionales.
Datos Referenciales	Son conjunto de valores o esquemas de clasificación que son referenciados por los sistemas, almacenes de datos, procesos y reportes, así como por los datos transaccionales y maestros. Ejemplos: países, instituciones, organismos, provincias, tipos de deudas, etc.
Metadatos	Literalmente significa datos acerca de los datos, describen o caracterizan otros datos para hacer más fácil su interpretación, recuperación y uso. Ejemplos: nombre de campos, longitud, tipos (metadatos técnicos), definiciones de campos, nombre de reportes, nombres de pantallas (metadatos del negocio).
Datos Históricos	Contienen hechos significantes que no pueden ser alterados excepto para corregir errores.
Datos Temporales	Mantenidos en memoria o tablas temporales para procesamiento rápido y propósitos técnicos.

Las especificaciones de los datos proveen la información necesaria para hacer, construir o producir algo.

Datos estándar: Son reglas que rigen cómo nombrar datos (tablas y campos), cómo definirlos, cómo establecer valor válido y cómo especificar sus reglas del negocio (English Larry, 2007).

Modelo de datos: Un modelo de datos según (Brodie.M., Mylopoulos, J. & Schmidt,J., 1984) es una colección de conceptos bien definidos matemáticamente que ayudan a

considerar y expresar las propiedades estáticas y dinámicas de una aplicación. Los términos más comúnmente utilizados son: el conceptual, el lógico y el físico, existiendo múltiples definiciones para cada uno (Hay, David C, 2003).

Reglas del negocio: Guías o principios que describen las interacciones del negocio (Ross, R.G., 2006) y establece reglas para acciones que dan como resultado el comportamiento y la integridad de los datos (Ross, R.G., 2005), (McGilvray, Danette, 2008). Existen varios tipos de reglas: de restricciones, guías, de computación, inferencia, de control de tiempo y disparadoras o triggers.

Los Metadatos (Hay, David C, 2006) y **Datos de Referencia** (Chisholm, M., 2001) también especifican los datos.

1.5 Metodologías y Marcos de trabajos de calidad de datos

La calidad de los datos es una disciplina diversa y esto se refleja en las metodologías y marcos de trabajo de calidad de datos actualmente disponibles (Dataqualitypro, 2010). Según Dylan (2009a), se necesita ser muy subjetivo para establecer qué componentes incluir. Si se desea construir una metodología de calidad de datos para un almacén de datos, en ese caso se pueden requerir las definiciones de la actividad justamente detalladas de cómo medir, evaluar y mejorar los datos a un nivel práctico; si se enfoca en la mejora de calidad de datos interna o es la meta para mejorar la calidad de los datos para la organización, puede haber mayores diferencias de la manera en que se agrupan los diferentes marcos de trabajo.

Se define una **Metodología de Calidad de Datos** (Batini, Carlo & Scannapieca, Monica, 2006) como un conjunto de guías y técnicas que, comenzando desde la información de entrada, define el proceso racional para usar la información para medir y mejorar la calidad de datos de una organización a través de las fases o etapas definidas y puntos de decisión. Diferentes tipos de conocimientos de entrada a la metodología en el caso más general es mostrada en la figura 1.2, las flechas representan la generalización jerárquica entre los conceptos: las colecciones de datos pueden ser los grupos internos o las fuentes externas y los grupos internos pueden ser flujos de los datos o bases de datos.

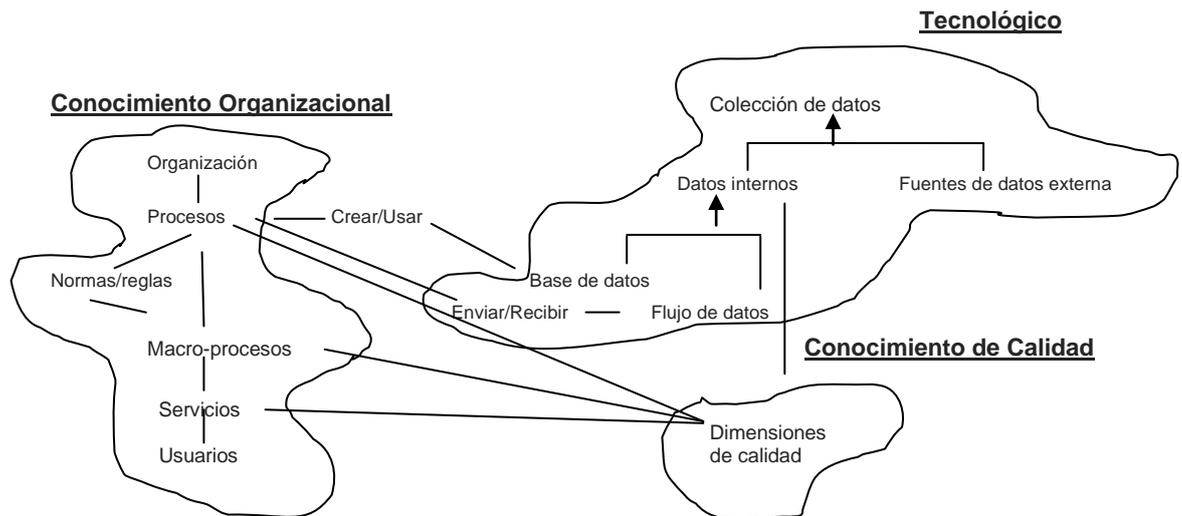


Figura 1.2 Conocimientos en la gestión de calidad de datos y mejoramiento de los procesos

Los tipos principales de conocimiento son:

1. La organización u organizaciones involucradas en los procesos, con las estructuras orgánicas relacionadas, funciones, normas y reglas.
2. Los procesos y macro-procesos realizados en la organización, que ejecutados juntos producen los servicios para los usuarios.
3. Los servicios entregados por los procesos y los usuarios que piden los servicios.
4. Las reglas de disciplina para la ejecución de procesos y macro-procesos.
5. La calidad de los procesos, macro-procesos y servicios, por ejemplo el tiempo de ejecución de un proceso, la utilidad de un servicio y la exactitud de información provista por el servicio.
6. Las colecciones de datos, correspondiendo a todas las bases de datos y flujos de los datos que son de interés a la organización. Para ejecutar los procesos, la organización tiene que guardar los datos permanentemente en el repositorio y para cooperar, ellos tienen que intercambiarse a través de los flujos de datos, ambos tipos de datos tienen que ser considerados partiendo de que los errores pueden ser afectados y propagados por ambos; en dependencia de su calidad, ellos influyen positivamente o negativamente en la calidad de los procesos.
7. Las fuentes de datos externas, a menudo más crítica que los datos internos para su calidad, no tiene ningún mando en los procesos.
8. Las dimensiones de calidad de datos influyen en la mejora de los procesos.

La calidad de datos es una disciplina diversa y esto es reflejado en las metodologías de calidad de datos actualmente disponibles en Data Quality Pro. Metodologías y frameworks no son las actividades "excepcionales" y deben ser una acción ininterrumpida para cualquier organización que estará aprendiendo, fallando, teniendo éxito y madurando con su enfoque para la calidad de datos constantemente, así que se debe ser muy subjetivo para determinar qué componentes incluir en la estructura de calidad de datos.

Una metodología de Calidad de Datos de propósito general (Batini, Carlo & Scannapieca, Monica, 2006) propone en la figura 1.3, las siguientes entradas y salidas en su estructura para los procesos de mejoras.

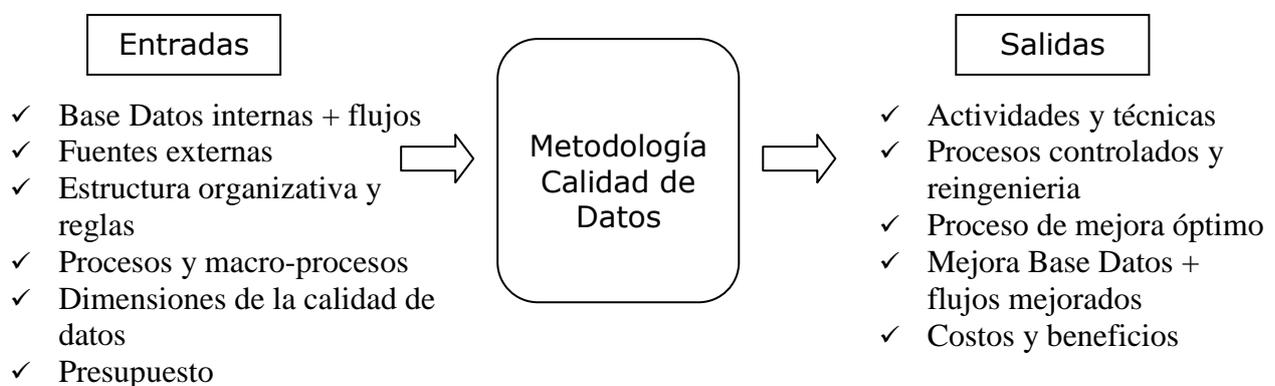


Figura 1.3 Entradas/Salidas de una Metodología de Calidad

A continuación se evalúan algunas de las metodologías y marcos de trabajo más representativos y de mayor impacto reportadas en la literatura.

1.5.1 TIQM: Gestión total de la calidad de la información

El marco general de TIQM (*Total Information Quality Management*) es exhaustivo y detallado, de impacto internacional, desarrollado por Larry English (1999) uno de los primeros pioneros de la gestión de calidad de la información y sigue siendo venerado como uno de los líderes más prominentes de la industria. TIQM es bien detallado en uno de los libros de calidad de información más populares de todos los tiempos del propio autor.

Se diseñó inicialmente para proyectos de almacén de los datos, pero su amplio alcance y nivel de detalles la caracterizan como una metodología de propósito general para todo tipo

de procesos de negocios habilitados para tecnología de la Calidad de la Información (CI) y cuenta con seis etapas: Evaluar definición de los datos y calidad de la arquitectura de información, Evaluar calidad de la información, Costos de medición y riesgos de la información de baja calidad, Mejora de información: reingeniería de datos y limpieza, Mejorar la calidad del proceso de la información: la prevención de defectos de datos y el Mantenimiento de un entorno informativo de calidad eficaz. Cada etapa implica las actividades que se asocian con la CI y cubre actividades tanto para el manejo de datos como de procesos. Es un sistema holístico de gestión total de información de calidad; significa que la calidad de información se ha convertido en una parte natural del comportamiento de la dirección y del personal.

También se describe cómo utilizar algunas de las herramientas de calidad y técnicas probadas de los sistemas de calidad, tales como: Deming (Deming, W. Edward, 2000) con los procesos PDCA (*Plan - Do - Check - Act*), Juran (Juran, 1995), (Redman, Thomas C., 2001b) y Six Sigma (Pande, Peter S., & Holpp Larry, 2002), (Gutiérrez, H. y De la Vara, R., 2004) con los procesos DMAIC (*Define - Measure - Analyze - Improve - Control*).

Un marco general de TIQM (English, Larry, 2003) es para todo tipo de procesos de negocios habilitados para tecnología de la información. Hai y Shouhong (2009) aseguran que la técnica del conjunto aproximado de reglas de inducción es una herramienta poderosa para TIQM y detallan las siete etapas de la producción de información:

Fuentes de información: Proporcionan datos para la producción de información.

Adquisición de información: Se logra mediante la captura de los datos de la organización. En esta etapa se describen las definiciones y los formatos de datos estructurados.

Almacenamiento de información: Son las bases y almacenes de datos. Más en general, los portales Web y repositorios de información no estructurados, son también almacenamiento.

Manipulación de la información: Es la etapa central de procesamiento de la información. Tiene dos componentes principales interrelacionados: el modelo y el procedimiento. Los modelos son las fórmulas y algoritmos de decisión, los procedimientos incluyen secuencias de flujo de trabajo, consultas y motores de búsqueda. Dimensiones tales como la precisión, acceso, oportunidad e integridad son particularmente relevantes para los procedimientos.

Presentación de la Información: La información se aporta a través de las interfaces de usuario. Dimensiones como interpretación, la facilidad de entender y la consistencia son particularmente relevantes para esta etapa de la producción de información.

Control de acceso: El usuario puede acceder a la información necesaria a través de la autorización y autenticación.

Utilización de la información: La información es un producto intangible que se produjo finalmente en la mente del usuario, es la etapa final de producción de información.

Recientemente, Larry English publicó la actualización "Calidad de la Información aplicada: Mejores prácticas para mejorar la información de empresas, procesos y sistemas", a su primer trabajo "Mejorando el almacén de datos y la Calidad de Información del negocio", creando algunos cambios a la metodología original y proporcionando varias secciones nuevas. En el nuevo trabajo aborda principalmente el uso de técnicas de probada eficacia de los sistemas de calidad para apoyar la Gestión de la Información (Dylan, Jones, 2009b). El cambio más significativo es el tratamiento más amplio de cómo establecer una cultura de información de calidad sostenible. Otro cambio importante, reside en la reestructuración del número de procesos del anterior TIQM.

A modo de resumen TIQM es la metodología de la década de los 90 que mayormente satisfizo a gerentes y empresarios. Mantiene elevadas indicaciones aplicando y generando consenso para la aplicación en una organización, también es excelente estableciendo las pautas para verificar la efectividad de mejoras y describe cómo utilizar algunas de las herramientas de calidad y técnicas probadas. Se considera como debilidad brindarle más importancia a la etapa TIQM 4 "Mejora de información del producto " que a TIQM 5, "Mejorar la calidad del proceso de la información", en realidad la prevención de defectos de datos es la competencia básica fundamental de cualquier sistema de calidad y no la corrección de datos que tiene un alto costo.

1.5.2 Conocimientos de Dirección de Empresa: El enfoque de Calidad de Datos

Enterprise Knowledge Management: The Data Quality Approach es un marco de trabajo de integridad de conocimientos, suministrado por David Loshin (2001b) y cubre componentes para desarrollar el *framework* económico que comienza por el valor de la calidad de datos,

procesos de control estadístico, la valoración del estado actual, el análisis de casos y cómo implementar reglas de la empresa para la calidad de datos. Escrito para directivos y técnicos, se enfatiza en los 17 pasos como muestra la tabla 1.5, para lanzar su iniciativa de calidad.

En el marco de trabajo se detallan los altos costos de datos malos y las opciones disponibles para transformar los datos en el verdadero conocimiento de una empresa, presentando información conceptual y práctica que complementa el interés en el almacenaje de los datos, minería de datos y descubrimiento de conocimiento. Provee un análisis de las dimensiones de la calidad de datos, y un marco para expresar reglas de calidad de datos, cómo medir esas reglas y evaluar niveles de la calidad de datos hacia una "Tarjeta de tanteo de calidad de datos". Esta tarjeta puede ser empleada como un punto de referencia y base para una mejora de calidad de la información.

Tabla 1.5 Pasos del framework de Conocimientos

<i>Pasos</i>
Paso 1. Reconocimiento del problema
Paso 2. Soporte de dirección y política de propiedad de datos
Paso 3. Difundir la palabra
Paso 4. Levantamiento de la cadena de información
Paso 5. Apuntes de la calidad de datos
Paso 6. Valoración del estado actual
Paso 7. Evaluación de los requerimientos
Paso 8. Seleccionar el Proyecto
Paso 9. Construir el equipo de trabajo
Paso 10. Construir el arsenal
Paso 11. Modelo del Metadato
Paso 12. Definir las reglas de la calidad de datos
Paso 13. Arqueología o minería de datos
Paso 14. Dirección de los proveedores
Paso 15. Ejecutando la mejora
Paso 16. Medir la Mejora
Paso 17. Construir en cada éxito

El marco contiene consejos de un consultor de calidad de datos muy exitoso, es el único libro en calidad de los datos que ofrece la perspicacia a gerentes y técnicos para atraer a profesionales en informática, sin embargo es demasiado extenso detallando las listas de casos a considerar para cada paso de la metodología. Cada caso se documenta muy bien lo que puede ocasionar perder de vista la metodología que está presentándose. Se notan analogías significantes a conceptos relacionados en la guía de calidad de datos de Redman (2001a), sobre todo el uso de procesos estadísticos para las medidas de calidad.

1.5.3 Metodología para la valoración de Calidad de Datos en los datos Financieros

Esta metodología es definida y aplicada en datos financieros (De Amicis, Fabricio & Batini, Carlo, 2004) en su trabajo: *Methodology for Data Quality Assessment on Financial Data*. Comprende cinco etapas que definen el reconocimiento y clasificación de variables primarias, análisis de técnicas y reglas de calidad de datos para la inspección de las dimensiones de calidad seleccionadas.

El objetivo de la metodología es la definición de la valoración objetiva cuantitativa y de la valoración subjetiva cualitativa de la calidad de datos en las variables de registro financiero. Se detalla más en dos perspectivas, la identificación y clasificación de variables y técnicas de análisis de datos; y la definición de índices apropiados, reglas de calidad de datos, medidas y estrategias para las valoraciones cuantitativas y cualitativas.

La figura 1.4 proporciona la descripción de las cinco fases de la metodología y de sus relaciones. Inicialmente, las variables financieras se seleccionan según su importancia y se clasifican para guiar los análisis adicionales y evaluar los resultados finales. La fase del análisis evalúa las dimensiones de calidad de datos (corrección semántica y sintáctica, completitud, consistencia interna y externa, unicidad y oportunidad) a través de métricas apropiadas y las reglas del negocio (propiedades semánticas no capturadas por las dimensiones previamente mencionadas) para identificar los errores de datos no estructurados y no controlados en los procesos de carga de datos y actualización. La etapa tercera comienza con las medidas de los errores proporcionando una valoración objetiva. La experiencia y profesionalidad es el recurso usado para realizar la valoración subjetiva cualitativa en cuarta etapa. El análisis comparativo entre las dos valoraciones se realiza en la etapa final.

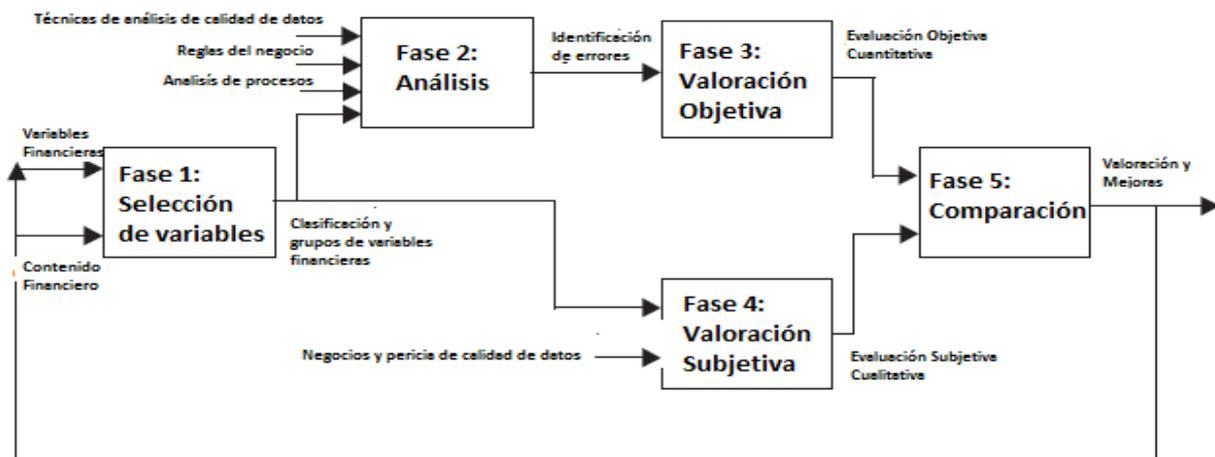


Figura 1.4 Etapas para la valoración de Calidad en los Datos Financieros

Se consideran dos aspectos a subrayar. Primero los datos financieros se refieren a variables que caracterizan los instrumentos financieros para realizar el negocio y pueden ser clasificados en cuatro categorías principales: datos del registro que describen los instrumentos financieros, datos diarios que se refieren a los precios y tipos de cambio, datos históricos mantenidos para relacionar las series en el tiempo, y los datos teóricos que corresponden al rendimiento de modelos financieros. Segundo la metodología está basada en la experiencia ganada en los proyectos de calidad desarrollados en diferentes bancos e instituciones financieras italianas y un banco suizo privado, aportando el uso de las técnicas de análisis disponibles en la literatura de manera entallada al dominio de los datos. Sin embargo no se mencionan soluciones o alternativas para corregir los errores y discrepancias detectadas entre las dos valoraciones (subjetivas y objetivas) como plan de mejoras para la calidad de datos financieros y prever errores futuros.

1.5.4 CDQM: Calidad de Datos - Conceptos, Metodologías y Técnicas

La "Metodología de calidad de datos completa" CDQM, es un *framework* (Batini, Carlo & Scannapieca, Monica, 2006) presentado en su libro: *Data Quality - Concepts, Methodologies and Techniques*. Esencialmente consta de tres fases: la reconstrucción, la valoración y la mejora. Uno de los aspectos obligatorios es el enfoque sobre los procesos del negocio y los servicios de la organización, esto ayuda a dirigir los esfuerzos en las áreas donde los gastos son más altos. Es una descripción extensa de técnicas que

constituyen el núcleo de la calidad de los datos, incluye el *matching*³ de datos, integración de los datos (Marotta, A. & Ruggia, R., 2005), la localización y corrección del error; tales técnicas se examinan en un marco metodológico comprensivo, también se analizan profundamente las definiciones de dimensión de calidad y los modelos adoptados. Proporciona con las metodologías de referencias, soluciones prácticas para las técnicas más eficaces, ejemplos de casos de estudio y se resaltan o discuten las diferencias entre las soluciones propuestas. Las fases y pasos de CDQM son mostrados en la tabla 1.6.

Tabla 1.6 Fases y pasos de la metodología CDQM

<i>Fases</i>	<i>Pasos</i>
<i>Fase 1: Reconstrucción</i>	<ol style="list-style-type: none"> 1. Reconstruir el estado y significado de las bases de datos más relevantes y flujos de los datos intercambiados entre las organizaciones. Construir la base de datos + las matrices de flujo de datos/organizaciones. 2. Reconstruir los más relevantes procesos del negocio ejecutados por las organizaciones y construir la matriz de los procesos/organizaciones. 3. Para cada proceso o grupo de procesos relacionados en un macro-proceso, reconstruir las normas y las reglas orgánicas que discipline el macro-proceso y el servicio proporcionado.
<i>Fase 2: La valoración</i>	<ol style="list-style-type: none"> 4. Chequear los mayores problemas relacionados con los servicios proporcionados a los usuarios internos y finales. Arreglar estas desventajas en términos de procesos y calidad de servicio, e identificar las causas de las desventajas debido a la baja calidad de los datos. 5. Identificar las dimensiones de calidad de datos (DQ) pertinentes y métricas, medir calidad de datos de las bases de datos y flujos de datos e identificar las áreas críticas.
<i>Fase 3: Seleccionar el proceso de mejora óptimo</i>	<ol style="list-style-type: none"> 6. Para cada base de datos y los flujos de datos, arreglar los nuevos niveles de DQ que mejoran la calidad del proceso y reducen los costos bajo el umbral requerido. 7. Concebir las actividades de re-ingeniería de procesos y escoger actividades de DQ que pueden mejorar el conjunto de objetivos del paso 6, relacionándolos en la matriz de datos/actividades para agrupar bases de datos y flujos de los datos involucrados en la mejora de DQ. 8. Escoger las técnicas óptimas para las actividades de DQ. 9. Conectar la matriz datos/actividades en la mejora de los procesos. 10. Para cada proceso de mejora computar costos aproximados y beneficios, escoger el óptimo verificando que en el equilibrio del costo-beneficio global se encuentra los objetivos del paso 6.

³ La palabra "*matching*" de origen inglés, se utiliza para la detección de duplicados y la relación entre dos fuentes de datos que no tienen campos de unión entre sí.

Es una metodología original, caracterizada por un equilibrio razonable entre la integridad y la viabilidad práctica de la calidad de los datos y el proceso de mejoras, trata todos los tipos de conocimiento: organizacional, tecnológico y de la calidad; por esta razón, se denomina Calidad de Datos Completa. No tiene en cuenta el ciclo de vida de la información.

1.5.5 Diez pasos hacia una Calidad de Datos y de confianza de Información

Ejecución de Proyectos de Calidad de los datos: Diez pasos hacia una calidad de datos y de confianza de información (McGilvray, Danette, 2008) detalla un *framework* de calidad de datos comprensivo, apuntado a la necesidad de cualquier organización de entender la logística de ejecutar iniciativas de calidad de datos bien estructuradas.

McGilvray presenta un enfoque sistemático y probado para mejorar la calidad de los datos mediante la combinación de un marco conceptual para la comprensión de la calidad de la información con las técnicas y las instrucciones para su mejoramiento (Dylan, Jones, 2009c). El enfoque se aplica a todos los tipos de datos y tipos de organizaciones, incluye ejemplos detallados y consejos prácticos para la ejecución de cada paso de la guía que se muestra en la figura 1.5; destaca que la calidad de los datos no puede sostenerse sin el gobierno de datos que proporciona caminos de interacción y comunicación para garantizar una representación adecuada del negocio, la información y la tecnología para tomar decisiones, identificar y resolver problemas e implementar cambios (Dylan, Jones, 2009d). Recomienda obtener la información fundamental necesaria para proceder con eficacia, pero no llegar demasiado lejos en los detalles que pueden ser innecesarios, siempre se puede volver atrás y obtener detalles adicionales si es necesario.

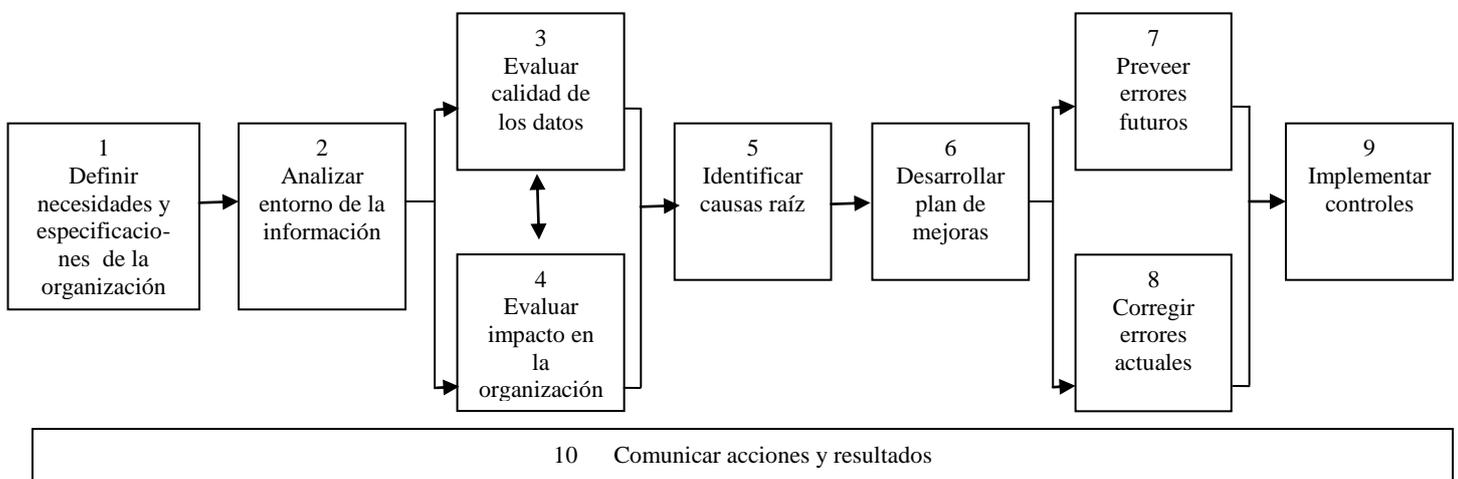


Figura 1.5 Actividades de los Diez Pasos

Algunas aproximaciones para construir el proyecto de calidad de datos son sugeridas por su propia autora. Las aproximaciones más comunes (McGilvray, Danette, 2008) son:

1. Establecer el caso del negocio u organización: Establecer el acercamiento puede ser una valoración exploratoria o una prueba rápida de la calidad de los datos evaluando un conjunto muy limitado de datos. Pasos recomendados: 1, 2, 3, 4, 10.
2. Establecer lo básico de la Calidad de Datos: Se aplica cuando la organización se ha comprometido a ir mejorando la calidad de los datos, hay apoyo y recursos; puede incluir la compra y/o empleo de herramientas de perfil de datos y limpieza. El objetivo no es solo descubrir los problemas sino identificar las causas raíz para elaborar un plan de acción de mejoras, requiere más nivel de detalles que la aproximación anterior. Pasos recomendados: 1, 2, 3, 4, 5, 6, 10.
3. Determinar las Causas Raíz: Se utiliza cuando se conocen muchos de los problemas de calidad de datos y se necesita una investigación más profunda, una vez que las causas son descubiertas se debe incluir las recomendaciones para implementar las mejoras. Pasos recomendados: 1, 2, 3, 4, 5, 6, 10.
4. Implementar mejoras: Ejecuta las recomendaciones desarrolladas cuando la valoración de calidad de datos y el análisis del impacto en la organización ha generado un plan para la calidad de los datos, tomará la acción para la corrección de errores y prevención futura. Pasos recomendados: resultados de la aproximación Básico de la Calidad de Datos adicionando los pasos 7, 8 y 10.
5. Direccionar la Calidad de Datos como un Individuo: Cualquier aproximación de las anteriores puede ser ejecutada. Se pueden encontrar algunas técnicas que formarán parte de los procesos individuales. Pasos recomendados: diversificar.
6. Integrar las actividades de Calidad de Datos en otros proyectos y metodologías: La mayoría de los proyectos como las aplicaciones de sistemas, migraciones de Planificación de Recursos de Empresa (ERP), almacenes de datos, proyectos de integración de datos; tienen beneficios de las actividades de calidad que incorporaron en sus planes. Un ejemplo es usar las dimensiones de calidad de datos durante la recolección de requisitos. Pasos recomendados: diversificar.

Esta propuesta permite la consulta rápida con un sencillo formato, destacando los conceptos claves y definiciones, los controles importantes, las actividades de comunicación y las mejores prácticas. Se considera un marco completo, práctico, contiene recursos de apoyo excelentes mediante un enfoque diferente; cubre el "Marco para la Calidad de la Información", "El ciclo de vida de la Información" y el "Ciclo de mejoramiento de Calidad de los datos". En un tema que es abundante en palabras y corto de consejos prácticos para su aplicación, Diez Pasos es una excepción refrescante.

1.5.6 Caracterización de Marcos de Trabajo y Metodologías

Existen otros trabajos cuyo principal objetivo es aportar metodología que permita la gestión de la calidad de los datos, y que no se mencionan de manera tan explícita por su menor influencia en nuestra investigación. En este grupo destacamos:

Data Governance Framework: Del Instituto de Gobierno de Datos. Aunque no es un *framework* de calidad de datos de por sí, obviamente lleva la calidad de datos al nivel de las empresas. Muchas organizaciones se están dando cuenta de la necesidad de implementar las estrategias de gobierno de datos para mantener sus iniciativas de calidad de datos.

New Zealand Ministry of Justice Data Quality Framework: Framework del Ministerio de Justicia de Nueva Zelanda: públicamente disponible, cubre el ciclo de vida de calidad de datos, las fases y las actividades (Definir - Medir - Analizar - Mejorar y Controlar), las dimensiones de calidad de datos, causas y costos de la calidad de datos, listas de verificación y la circulación de procesos de valoración. Un buen ejemplo de cómo un *framework* bien estructurado debe ser formulado basado en el ciclo DMAIC.

TDQM: Es un esfuerzo conjunto entre miembros del programa de *Total Data Quality Methodology* (TDQM), el programa de calidad de información de *Massachusetts Institute of Technology* (MIT) en Boston, CITM en Berkeley, agencias gubernamentales como la Marina de los EE.UU y socios de la industria. Proviene de la metodología de Gestión de la Calidad Total (TQM) que se desarrolló en el sector de fabricación. Fue ampliamente difundida en la década del 90 en los EE.UU y luego extendida a otros países. Contiene una guía para encontrar los errores, tiene un acercamiento claro orientado para el manejo de procesos.

De manera general, resumiendo todos los acercamientos a las metodologías y *frameworks* evaluados, a pesar de su evolución aún presentan carencias, ya que algunos no incorporan el plan de mejoras como exigencia para analizar y especificar qué debe hacerse, no tienen en cuenta la inevitable interacción con el ciclo de vida de la información, o no muestran alternativas que combinen varias técnicas en la prevención y solución de problemas. Podemos establecer que las fases comunes básicas (Dylan, Jones, 2009a) se distinguen entre la valoración y mejora de los procesos.

Las fases para el proceso de valoración son:

- Análisis de los datos, que colecciona el conocimiento de los datos, su arquitectura, los flujos y reglas de dirección de datos (Scofield, 2008). Esto puede lograrse examinando la documentación de los datos y los esquemas lógicos o a través de las entrevistas.
- Análisis de requisitos, que colecciona las sugerencias generales de los usuarios y gerentes ante las posibles causas de errores de los datos, determina los objetivos futuros para lograr para la calidad de los datos.
- Buscar áreas críticas, que selecciona lo más relevante en las bases de datos y los flujos de datos para ser analizado en detalle.
- Modelos de procesos, que describe el proceso (o procesos) según un modelo formal o semi-formal.
- Establecer las dimensiones de calidad y realizar las medidas en la base de datos, o si es impracticable o demasiado caro, hacerlo en una muestra.
- Costo de evaluación de la baja calidad, que estima los costos de los procesos debido a la pobre calidad de los datos.
- Evaluación de los beneficios, que estima el aumento de los ingresos y/o nuevos beneficios intangibles que derivan del posible aumento de calidad de los datos.
- Asignación de responsabilidades en los procesos, que determina responsabilidades por las actividades de obtención de datos.
- Asignación de responsabilidades en los datos, que asigna responsabilidades en el control

de los datos.

- Selección de las herramientas y técnicas más convenientes y disponibles para el contexto organizacional (WILSHIRE & DAMA, 2007), el conocimiento del dominio disponible y el presupuesto.

Las fases comunes para el proceso de mejora son:

- Encontrar las causas de errores (Maydanchik, 2007) para analizar las posibles causas de la desviación de las dimensiones de calidad de los valores designados.
- Diseñar la solución de mejora que seleccione entre las actividades de calidad las más eficaces a ser realizadas para producir los objetivos (Uribe, 2010).
- Establecer los procesos de control en la producción de los datos que permitan supervisar y restaurar la calidad deseada en las dimensiones durante la ejecución del proceso.
- Diseñar la solución de mejora en los procesos controlando las actividades ideadas en la fase de valoración y encontrar los pasos que producen las mejoras correspondientes del actual proceso.
- Rediseñar procesos, encontrar los cambios radicales que lleve a la mejora de calidad de datos de modo significativo.
- Manejar la solución de mejora bajo la perspectiva de encontrar las nuevas reglas de la organización para la calidad de los datos.
- Chequear la efectividad de las mejoras, que establece evaluaciones periódicas y supervisión de actividades (Scofield, 2008) que proporcionan la regeneración del proceso y ponen a punto su condición dinámica.

1.5.7 Herramientas y productos líderes en la Calidad e Integración de los datos

Muchos esfuerzos se han realizado desde la academia y desde la industria para generar conocimiento y herramientas relacionadas a la gestión de calidad de datos. Para la búsqueda de una herramienta, la elección dependerá del entorno y las necesidades. Algunas herramientas se especializan en determinados tipos de datos, como direcciones o cuentas de clientes, otras proveen entornos para que el usuario pueda definir sus propias

reglas. La tendencia en estos casos, es ofrecer soluciones integradas de calidad de datos junto con soluciones para federación, integración, procesos de ETL (Extracción, Transformación y Carga de datos) y MDM (*Master Data Management*). Muchas de las grandes empresas han promocionado soluciones integradas de este tipo, incluyendo a Microsoft, IBM y Oracle.

Según Loshin, “Las herramientas de calidad de datos y la tecnología son necesarias para apoyar el análisis y la valoración de las fases”. La gestión de la calidad de datos incorpora un ciclo eficaz mostrado en la figura 1.7, consistiendo esencialmente en dos fases: el análisis y valoración, la supervisión y la mejora (Loshin, David, 2001c, 2009a), (Dylan, Jones, 2009e).

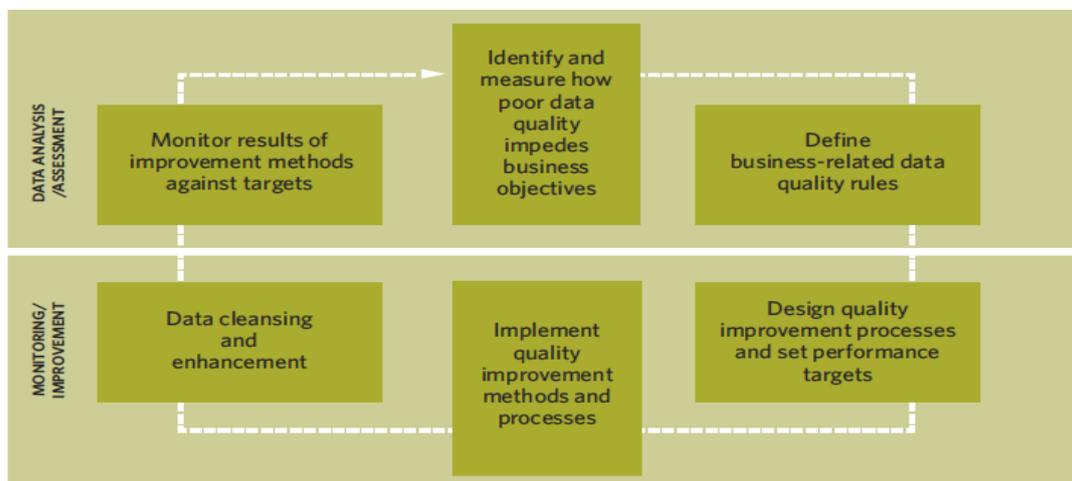


Figura 1.7 Ciclo de la gestión de calidad. Tomado de SearchDataManagement.com

Por otra parte, la necesidad de introducir una oportunidad para la gestión integrada de la información organizacional que lleve a cabo políticas, procedimientos y la infraestructura, compartiendo el uso exacto, oportuno, consistente y completo del dominio de los datos, es el desafío de lo que es conocido como la gestión de datos maestros (MDM), integrando herramientas, personas y prácticas para organizar la información importante de la organización y gobernar su calidad, utilizando una vista unificada de la información y sincronizada para lograr los objetivos (Loshin, David, 2009b).

Este enfoque ofrece numerosas ventajas. Los cimientos de integración de datos permiten la identificación, acceso, transformación, reconciliación semántica, gestión y suministro de

datos empresariales de forma segura, consistente y oportuna. Otra ventaja clave es la capacidad para aprovechar un entorno único en la gestión de los cambios y el análisis del impacto a través de los metadatos. Finalmente, una infraestructura de integración de datos es capaz de apoyar a los usuarios que requieran los datos con diferentes niveles de latencia o casi en tiempo real.

Los productos líderes en la Calidad e Integración de los datos son mostrados en el Anexo 2 y Anexo 3 respectivamente.

1.6 Conclusiones parciales

La necesidad de definir y elaborar el procedimiento que se propone resulta de cualquier esfuerzo que haga uso de una metodología a adoptar, la parte más difícil será realizar la mejor elección de actividades o pasos y cómo direccionarlos. La clave está en conocer cuáles son los requerimientos, qué necesitamos hacer, qué tiempo nos llevará y cómo llevarlo a cabo. Un proyecto puede ser (1) enfocado en la mejora de calidad de los datos, (2) una consciente aplicación de pasos específicos o técnicas de una metodología para resolver un problema dentro del área de responsabilidad, o (3) las tareas de calidad de los datos integradas en otros proyectos o metodologías; como el desarrollo de una nueva aplicación, la migración o integración de datos en un Almacén del datos o ERP.

Después de haber realizado un estudio acerca del estado actual de los marcos de trabajo y metodologías de Calidad de Datos más representativos que se abordan en la literatura, *Diez pasos hacia una Calidad de Datos y de confianza de Información*, es una de las publicaciones, hasta la fecha, más eficaces disponibles en la calidad de los datos que permite la mejor adopción con su enfoque sistemático y probado, mediante la combinación de un marco conceptual con las técnicas e instrucciones para el mejoramiento.

Realmente no es ineludible utilizar todos los pasos, el procedimiento será estructurado de manera que determine los objetivos propuestos con el propósito de elaborar un proyecto que cumpla con el segundo enfoque mencionado (2) y la ejecución de los pasos asociados a la aproximación para Determinar las Causas Raíz, siguiendo la recomendación de Larry English en entrevista con *Data Quality Pro*: “Una organización nunca debe corregir los datos antes de identificar la causa de la raíz de los defectos de la información; y definir el

mejoramiento de los procesos que eliminan las causas de procesos defectuosos” (Dylan, Jones, 2009b).

CAPÍTULO II. PROCEDIMIENTO PARA LA GESTIÓN DE LA CALIDAD DE DATOS EN EL SIB

2.1 Evaluación del modelo general del actual sistema SIB

El BCC⁴ para asumir todas sus funciones necesitó de un sistema operacional dinámico y parametrizable. Se definió una base de datos central con facilidades de mantenimiento y extensión, con módulos para la captación, almacenamiento y recuperación de la información en función del mecanismo de seguridad implementado para el acceso a cada módulo. Sin embargo, tiene inconvenientes por la necesidad de aplicar transformaciones a los datos origen para su almacenamiento, por tal motivo la recuperación de la información, es a su vez un problema en cuanto a tiempo de espera y el volumen que se quiere consultar.

Toda la información se almacena reducida al concepto de Indicador Económico, entendiéndose por tal, el valor de un parámetro que refleja un aspecto económico o financiero cualquiera. Para que el Indicador Económico sea útil, tiene que ser ubicado de alguna forma en tiempo y espacio. La ubicación en el espacio se realizó utilizando diferentes criterios o combinaciones de estos: geográfico, político - administrativo, institucional, o ramal. La ubicación en el tiempo de un indicador puede referirse a una fecha puntual ó a un período de tiempo.

El almacenamiento físico de estos indicadores (Guerrero, Julio, 2006) se registra en tablas transaccionales (denominadas Maestros) que describen la ubicación en el espacio, el concepto, el valor y el tiempo del indicador en cuestión. Consultar esquema del modelo lógico en el Anexo 4.

La arquitectura modular del sistema operacional como se observa en la figura 2.1, facilita la personalización y adaptabilidad por un lado y por otro la facilidad de integración. Cada módulo gestiona funciones diferentes pero se interrelacionan entre sí compartiendo información mediante la base de datos centralizada que contiene un volumen grande de datos (datos históricos) provenientes de fuentes heterogéneas mediante procesos de descargas de datos complejos.

⁴ En este trabajo utilizaremos indistintamente Banco Central de Cuba y "organización", asumiéndolas como equivalentes

La evaluación de la calidad de una base de datos financieros íntegramente es una tarea difícil y costosa que raramente se hace, adicionalmente, no existe una metodología completa para la evaluación de datos financieros que haya sido propuesta en alguna literatura relevante (De Amicis, Fabricio & Batini, Carlo, 2004).

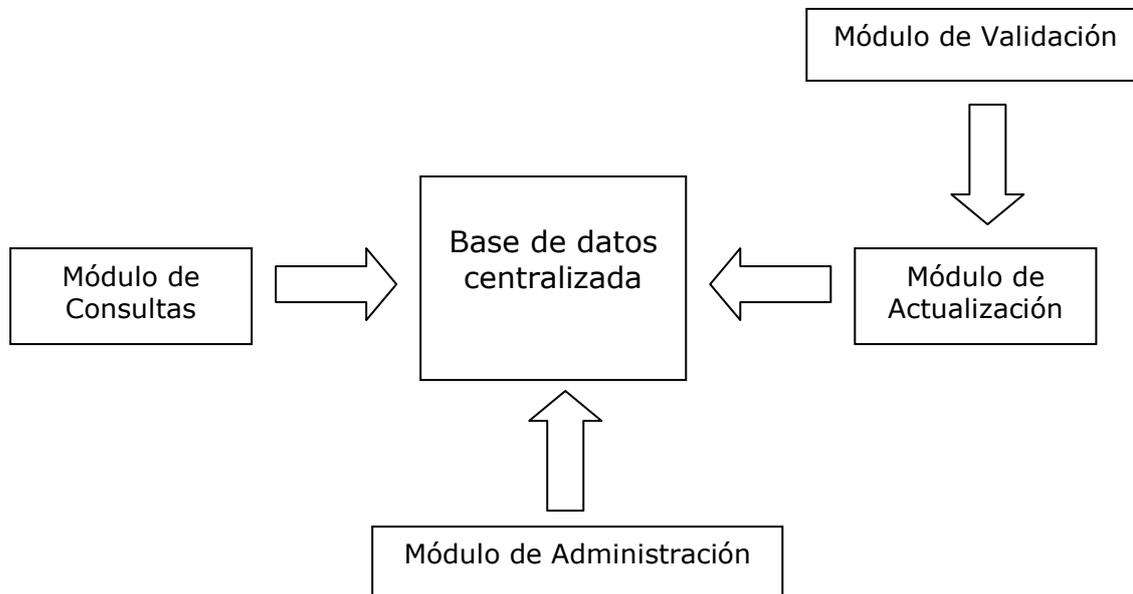


Figura 2.1 Esquema de la arquitectura modular del SIB

Uno de los principios de gestión de la calidad es el enfoque basado en hechos para la toma de decisión, el cual plantea que las decisiones eficaces se basan en el análisis de los datos y la información. Para garantizar estos aspectos se debe hacer, en principio, un diseño apropiado de la base, tabla o lista de datos, con el fin de definir correctamente los atributos o tipos de datos en la misma; y posteriormente realizar un adecuado diseño de los procesos de producción de datos, garantizando que los datos lleguen a la base o tabla de datos, libres de defectos y con las demás características deseadas. Para mejorar definitivamente la calidad de los datos, la tecnología informática no es la única solución, ya que para encontrar las causas de los problemas e incluso para mejorarlos, se necesita enfocarse en cuestiones no sólo relacionadas con el uso de tecnología de punta (Heredia, J. y Vilalta J., 2009). Pretender mejorar la calidad sólo a partir de buscar y corregir errores sería actuar sobre los efectos del problema, cuando realmente lo más importante es detectar las causas de los errores de calidad.

2.2 Mejores prácticas y pautas para la estructuración del procedimiento

En la figura 2.2 se representan las etapas del procedimiento en una secuencia lógica de pasos. Ha sido construido a partir de las mejores prácticas y pautas reconocidas en la metodología de referencia (McGilvray, Danette, 2008) y en cada uno de los pasos se examinan diferentes aspectos y resultados dependiendo de los requerimientos en el tiempo. Las instrucciones, técnicas, ejemplos y plantillas representan la conducta correcta para comprender y decidir qué es lo relevante y apropiado en cada paso.

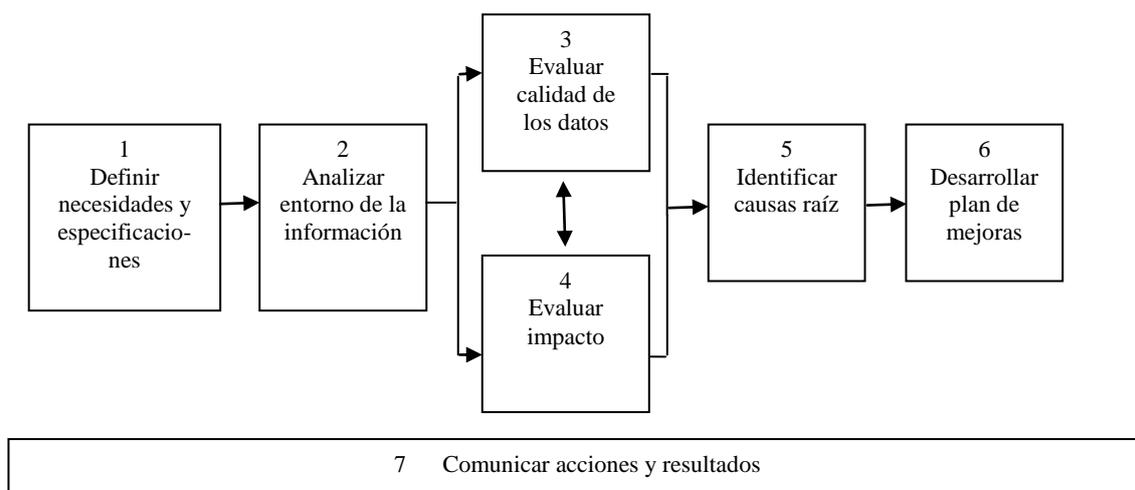


Figura 2.2 Pasos que constituyen el procedimiento en el SIB

La construcción del procedimiento se realiza sobre las siguientes premisas: **seleccionar** precavidamente de cada paso las actividades que apliquen a la situación del BCC. El **nivel de detalles** variará en cada uno de los pasos de acuerdo a las necesidades, comenzando por el nivel más alto y trabajando a nivel del detalle si es útil; será **reiterativo** si información adicional es descubierta a lo largo del proceso y las conclusiones más tempranas pueden necesitar ser revisadas nuevamente; emplearlo **apropiadamente** de manera que las actividades puedan ser aplicadas de conjunto con otros proyectos como la integración o migración de los datos.

Simultáneamente se sustenta en los principios: **escalable** para que pueda ser utilizado desde una persona en pocas semanas, hasta un equipo de proyecto durante varios meses dependiendo del tiempo y recursos requeridos; **reusable** para brindar el conocimiento adquirido de forma tal que sea mejor entendido y guíe a tomar mejores decisiones, asegurando que los descubrimientos, documentación o actualizaciones estén disponibles;

flexible permitiendo la habilidad de tratar con el cambio continuo; **enfocado en procesos** analizando por qué existen los problemas y no quién los está creando; **mejoramiento continuo** por su propia naturaleza iterativa.

Una de las características del procedimiento es su flexibilidad, su adopción requiere el buen juicio, conocimiento del negocio y creatividad. En cada paso del proyecto a lo largo de todo el ciclo de vida, como aparece en la figura 2.3, se describen conocimientos y procesos; a su vez, cada proceso se describe en términos de entradas, salidas, herramientas y técnicas. Las entradas y salidas (ficheros y documentos), las herramientas y técnicas son mecanismos aplicados a las entradas para crear salidas. Uno de los aspectos más importantes para la exitosa implantación de una metodología para la gestión de calidad de los datos es la planificación (Montero, L., 2004).

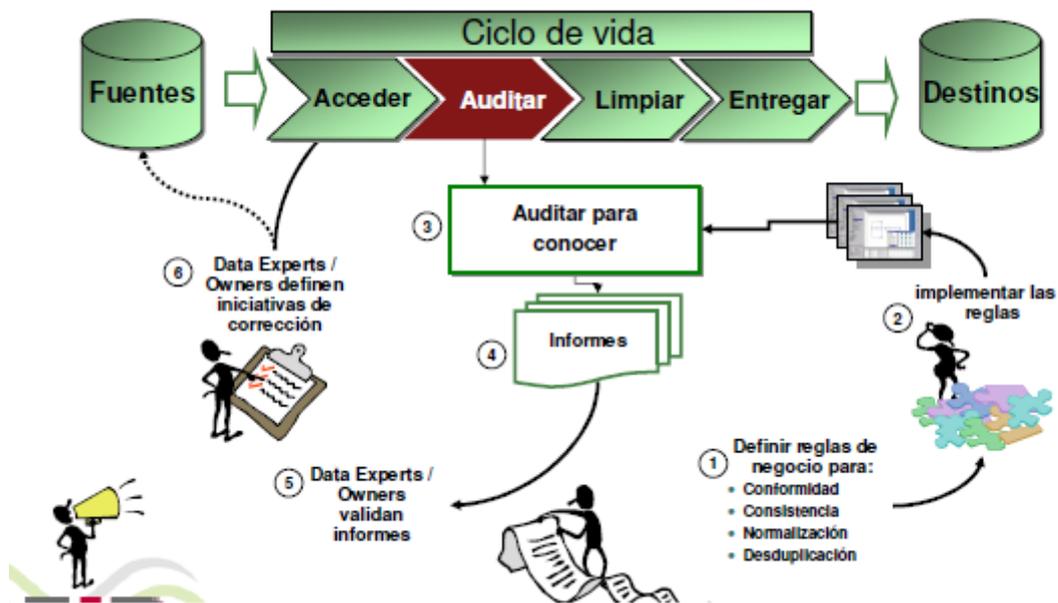


Figura 2.3 Ciclo de vida del procedimiento de calidad de datos

2.3 Descripción de los pasos o etapas del procedimiento

Su concepción teórica incluye lo relevante de las mejores prácticas y guías para concebir la calidad de datos en el SIB. Es un procedimiento iterativo e incremental por la propia naturaleza del proceso. Está conformado por un conjunto de etapas y actividades de manera coherente y organizada, ha sido diseñado de forma clara y sencilla para la comprensión y aplicación práctica de los analistas y expertos de la DEMF.

<i>Etapas o Pasos</i>	<i>Descripción de las actividades</i>
Paso 1. Definir necesidades y especificaciones	Definir los objetivos y estrategias, identificar el problema claramente, establecer la oportunidad o la meta para guiar el trabajo a lo largo del proyecto.
Paso 2. Analizar entorno de la información	Compilar y analizar la información sobre la situación actual y el medio ambiente de la información. Diseñar el método (herramienta) de captura de los datos y el plan de valoración de los mismos.
Paso 3. Evaluar calidad de los datos	Evaluar la calidad para las dimensiones de calidad de datos aplicable al SIB. Los resultados de valoración proporcionan la base para los pasos del futuro, como identificar las causas raíz y necesidad de las correcciones.
Paso 4. Evaluar impacto	Utilizar variedad de técnicas para determinar el impacto de la pobre-calidad de los datos. Esto proporciona la aceptación para establecer la mejora, para ganar el apoyo en la calidad de información y para determinar las inversiones apropiadas.
Paso 5. Identificar las causas raíces	Identificar y priorizar las verdaderas causas de los problemas de calidad y desarrollar las recomendaciones específicas para direccionarlas.
Paso 6. Desarrollar plan de mejoras	Desarrollar las recomendaciones específicas para la acción y ejecutar planes de mejoras.
Paso 7. Comunicar acciones y resultados	Documentar y comunicar los resultados de la prueba de calidad, las mejoras realizadas y los resultados de esas mejoras. La comunicación es parte importante en cada paso.

2.3.1 Etapa 1. Definir necesidades y especificaciones

Desarrollado de acuerdo con los interesados (*stakeholders*) o analistas de la DEMF para el entendimiento del problema, así como el alcance del estado actual de almacenamiento de los datos, comprender el contexto del negocio con los procesos internos y externos (Eppler, 2003), las personas/organizaciones involucradas, entender el sistema operacional implementado con sus funcionalidades y capacidades, fijar las normas de seguridad y definir el conjunto de datos y relaciones a valorar.

En esta etapa es importante que los objetivos de la dirección y estrategias conduzcan a todas las acciones y decisiones correctas. La decisión de llevar a cabo el proyecto comienza siempre con una pregunta, ¿Por qué es importante para la organización?, independiente de los elementos que puedan obtenerse, a favor o en contra, en las actividades subsiguientes de este procedimiento.

Es imprescindible la constitución del equipo de trabajo para el desarrollo de las actividades, por cuanto las características de las relaciones humanas entre los miembros seleccionados puede ser un factor clave del éxito o fracaso del proyecto. El equipo de trabajo debe estar integrado por: el director del departamento al tener conocimientos de las necesidades, recursos disponibles y estar investido con poder de decisión, los analistas o expertos del negocio que deben poseer conocimiento profundo del dominio del problema y las necesidades, especialistas informáticos que tengan experiencia en el dominio de la solución al problema planteado y que posean los conocimientos técnicos especializados imprescindibles y los usuarios finales que posean conocimiento funcional profundo del dominio del problema y del entorno sobre los que incidirá la aplicación del procedimiento a desarrollar. Para que funcione adecuadamente la comunicación y fluya coherentemente el proceso, debe tenerse en cuenta que un jefe de proyecto debe ser designado con la preparación técnica (en informática) adecuada al alcance del procedimiento a desarrollar y con conocimiento razonable de la problemática funcional del sistema y su entorno. El seleccionado debe poseer las particulares, habilidades y cualidades imprescindibles para ser líder de un equipo de trabajo de estas características.

Al inicio se recaba la **recogida de información** de directores y gerentes de los departamentos del BCC (directivos de nivel alto y medio), porque son los que cuentan con una visión lo suficientemente general del problema como para ayudar a crear una visión lo adecuadamente global del proyecto; después en el análisis detallado del problema se debe acudir a fuentes más específicas, como por ejemplo, los usuarios finales para analizar y evaluar conjuntos interesantes de datos y sus relaciones con el objetivo de entenderlos y formalizar criterios de calidad, determinar los valores ausentes (Tan, 2006, pp. 40-41), (Acuña, E. & Rodríguez, C., 2009), valores atípicos (Barnett, V. & Lewis, T., 1994) o aceptables y óptimos según las necesidades de la DEMF; así como identificar, evaluar la entrada y el almacenamiento de los datos para ajustar/recomendar procesos que impidan datos inesperados en el repositorio.

Las técnicas de recogida de información que pueden ser utilizadas son:

Entrevista: resulta una técnica muy aceptada y su uso está ampliamente extendido. Permite tomar conocimiento del problema y comprender los objetivos de la solución buscada. A través de esta técnica el equipo de trabajo se acerca al problema de una forma natural.

Brainstorming (Tormenta de ideas): es también una técnica de reuniones en grupo cuyo objetivo es coleccionar ideas individuales, presentarlas y explorarlas de manera ordenada en un ambiente libre de crítica.

Cuestionarios y listas de chequeo: esta técnica requiere que el analista conozca el ámbito del problema en el que está trabajando. Consiste en redactar un documento con preguntas (*checklist*) cuyas respuestas sean cortas y concretas, o incluso cerradas por unas cuantas opciones en el propio cuestionario. Este cuestionario será cumplimentado por el grupo de personas entrevistadas o simplemente para recoger información en forma independiente de una entrevista.

Plantillas: esta técnica recomendada por varios autores (Escalona, M.J., Torres, J., & Mejías, M., 2002; McGilvray, Danette, 2008) tiene por objetivo describir los requisitos mediante el lenguaje natural pero de una forma estructurada. Una plantilla es una tabla con una serie de campos y una estructura predefinida. Las plantillas eliminan parte de la ambigüedad del lenguaje natural al estructurar la información.

Como principio, es necesario exponer la interacción entre las fases del ciclo de vida de la información y los componentes que impactan en la calidad de los datos para evaluar los problemas y el entorno de la información. Obsérvese figura 2.4.

	Planificar	Obtener	Almacenar	Mantener	Aplicar	Disponer
Datos (Qué)	- Objetivos de la organización - Qué datos soporta y necesita - Reglas y datos estándar	- Datos que entran en el sistema operacional	- Datos almacenados y compartidos	- Datos almacenados y compartidos - Datos transformados	- Información disponible para su uso y automatizar procesos	- Datos archivados o eliminados
Procesos (Cómo)	- Qué son los procesos de alto nivel	- Fuentes de obtención de datos(internas/ externas) - Entrada de datos al sistema	-Qué son los procesos para almacenar y compartir datos	- Cómo los datos son actualizados y monitoreados para detectar cambios - Triggers para mantenimiento	- Cómo los datos son usados - Cómo la información es accedida , asegurada y disponible	- Cómo los datos son archivados o eliminados -Triggers para archivar
Personas/ Organización (Quiénes)	- Quiénes identifican objetivos, prioridades y recursos - Quiénes desarrollan reglas, procesos, estándares	- Quiénes adquieren información de las fuentes - Quiénes entran nuevos datos en el sistema	- Quién apoya la tecnología para almacenar y compartir los datos - Quiénes están involucrados en esta fase	- Quién decide que puede ser actualizado o cambiado en el sistema - Quién es responsable de la calidad y necesita conocer los cambios	- Quiénes directamente acceden a los datos - Quién usa información	- Quién archiva o elimina datos - Quién necesita estar informado
Tecnología (Cómo)	- Nivel de arquitectura y la tecnología soportada	- La aplicación crea nuevos registros en el sistema	- Tecnología para almacenar y archivar datos	- Cómo los datos son actualizados y mantenidos en el sistema	- Cómo las reglas del negocio son aplicadas en la arquitectura	- Cómo los registros e información son eliminados y/o archivados desde el sistema

Figura 2.4. Matriz de interacción - Fases del ciclo y componentes

Aproximación a los problemas y necesidades del BCC

Para la DEMF, dirigir datos como una fuente de información estratégica e invertir en su calidad (Trillium, Software, 2004a) es tomar la delantera en relación con la reputación y eficiencia de aquellos que lo dejan de hacer, garantizando al BCC la información estadística de calidad necesaria y verificando que cada una de las siguientes problemáticas sea considerada:

- ✓ Fuente externa de datos segura: Verificar el origen seguro de los datos de los contribuyentes (OACEs y SBFN).
- ✓ Costos innecesarios o excesivos: gasto de tiempo y corrección de datos una vez incorporados al repositorio.

- ✓ Incrementos de riesgos: que los datos puedan ser expuestos sin previo análisis de errores, imprecisiones o inconsistencias.
- ✓ Datos y procesos compartidos: varios procesos u otros sistemas informáticos (Sistema de Circulación Monetaria, Sistema de Licencias, CIR-Central de Riesgo, SAB-Sistema Automatizado de Balances, SRO-Sistema de Riesgo Operacional, etc.) que se nutren del repositorio de datos pueden ser impactados por los problemas de calidad.

Se utilizará la siguiente plantilla para categorizar la aproximación a los problemas:

Plantilla 2.1. Bases del Problema

Aproximación del Problema						
No.	Problema	Datos	Procesos	Personas/ Organización	Herramienta/ Tecnología	Comentarios

Con la información elemental recopilada, pero no trivial, se elabora el Informe preliminar de necesidades, que constituye la base para el **estudio de viabilidad del proyecto**. El estudio de factibilidad es una actividad compleja que requiere experiencia y permite realizar el análisis y la evaluación de las alternativas que se generan desde los puntos de vista siguientes:

Económico: ¿vale la pena invertir en el proyecto? ¿Los beneficios compensan los costos?,

Técnico: ¿se encuentra disponible la tecnología necesaria?,

Legal: se debe dilucidar si los objetivos y estrategias atentan contra alguna ley o procedimiento,

Operativo: se debe determinar si el procedimiento se puede evaluar de manera efectiva en la organización. ¿Encaja en la filosofía del BCC?

Posteriormente se selecciona la alternativa que permita resolver mejor el problema planteado y se realiza la argumentación detallada que debe comprobar: la contribución a los objetivos de la organización, la posibilidad de implementar con la tecnología actual y la integración con los procesos actuales de acuerdo al grado de necesidad.

Como mecanismo de evaluación y control se puede utilizar una lista de comprobación (*checklist*) que responda, entre otras, las preguntas siguientes:

¿Los problemas y objetivos del BCC están bien definidos y argumentados?

¿Los procesos de alto nivel (validación, actualización, administración), la manipulación de los datos, las personas, departamentos y tributantes involucrados, así como la tecnología; están correctamente asociados con el problema y documentados?

¿Se cuenta con el equipo de trabajo apropiado, junto con los documentos legales y procedimientos aplicables?

¿Se tiene un plan de comunicación inicial creado?

¿Existe otra comunicación necesaria en este momento que deba ser completada?

La salida de las actividades de esta etapa es una declaración informal de requisitos (escrita) y la aproximación al problema, que constituye a su vez la entrada a la siguiente etapa.

2.3.2 Etapa 2. Analizar entorno de la información

Conocer el entorno o ambiente, figura 2.5, permite el mejor análisis y toma de decisiones para continuar el proyecto, asegurar que la valoración de los datos está asociada con los problemas detectados, entender y comparar requerimientos y especificaciones contra datos de calidad, conocer datos relevantes, procesos, personas/organizaciones y sus roles, tecnología, documentar el ciclo de vida mediante sus fases y finalmente elaborar un plan realista de captura y evaluación de los datos sobre la base de tres reflexiones importantes: ¿Qué es relevante?, ¿Qué es apropiado?, ¿Cuál es el nivel de detalle adecuado?.

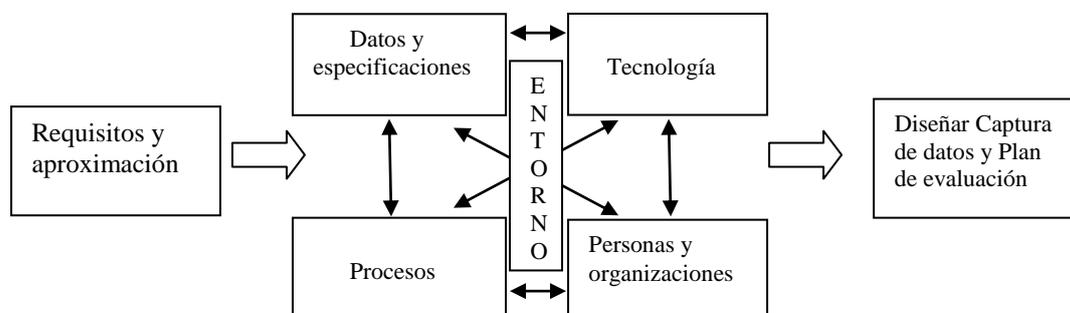


Figura 2.5. Flujo de las actividades – Entorno de la información

Datos y especificaciones

1. Los datos deben ser descritos desde un más alto nivel por los términos del negocio hasta los detalles de las actuales tablas y campos donde el dato es almacenado, siendo necesario capturar el lenguaje utilizado a través de entrevistas, evaluación de documentos (en este caso la metodología asociada a cada modelo en cuestión) y las pantallas del actual sistema informático.
2. Comprender y documentar el modelo de datos analizando su esquema conceptual, lógico y físico, así como las entidades y sus relaciones.
3. Crear una lista detallada para los datos interesantes como se propone en la plantilla 2.2, recolectando las especificaciones para cada uno de los atributos de interés, asociando las reglas, datos estándares o referenciales, metadatos, información estructural dentro de la base de datos (llave primaria o foránea, restricciones, disparadores o procedimientos almacenados para validaciones y forzar reglas) u otra información pertinente agregándola a la plantilla.

Plantilla 2.2. Lista detallada del repertorio de datos

Aplicación/ base de datos: SIB								
Dato Categoría	Tabla	Nombre del Atributo	Tipo de	Tamaño/ Longitud	Formato	Descripción	Dominio	Otros

Tecnología

1. Comprender el entorno que puede ser desde un alto nivel técnico (repositorio) hasta un nivel bajo. Conocer el funcionamiento del actual sistema operacional, cómo manipula los datos y los procesos involucrados o utilizados por las personas.
2. Hacer uso de la documentación existente y conocimientos de la Información Tecnológica (IT) para relacionar los recursos con el ciclo de vida de los datos (Planificar, Obtener, Almacenar, Aplicar y Disponer)
3. Determinar cualquier impacto en los datos a través de la tecnología: mover datos entre bases o tablas, transformaciones para alcanzar algún formato requerido (Tierstein, 2005).

Procesos

1. Identificar los procesos con un nivel apropiado de detalles. Comenzar con las funciones de la dirección o de las áreas responsables, listar los procedimientos y actividades que se ejecutan para cumplir estas funciones. Obsérvese la figura 2.6.
2. Relacionar datos y procesos relevantes e indicar las fases del ciclo de vida que tienen lugar.

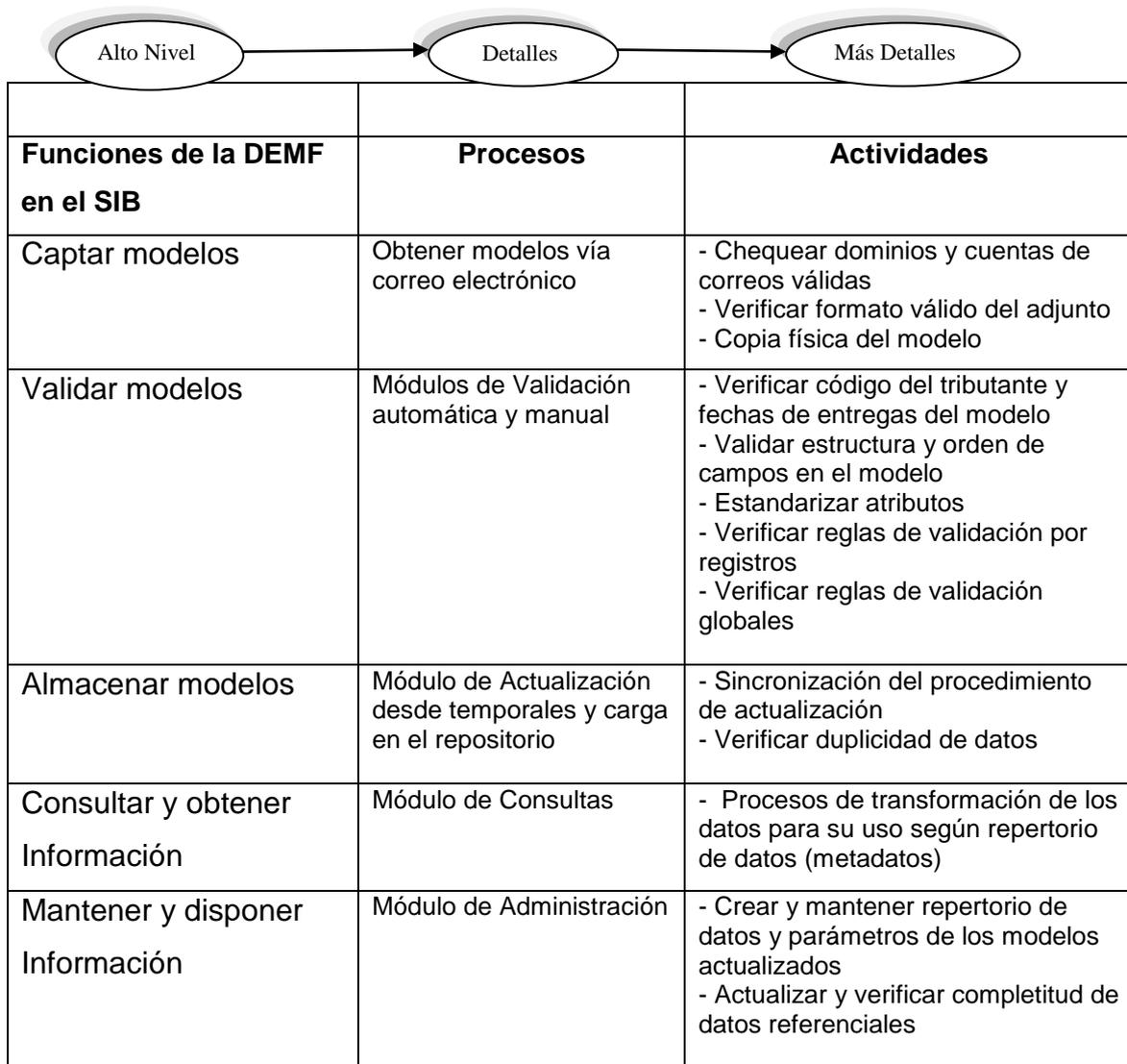


Figura 2.6. Niveles de detalles de los procesos

Personas y organizaciones

1. Identificar las instituciones tributantes, los grupos de personas y departamentos; examinando los roles y las responsabilidades apropiadas.

2. Relacionar los roles con las fases del ciclo de vida de los datos y comprender quiénes proveen la entrada e impactan los datos. La siguiente tabla describe conceptos acerca de roles que afectan la calidad de la información.

Tabla 2.1 Roles y Fases del ciclo de la información

<i>Roles</i>	<i>Fase del ciclo de vida de la Información</i>
<p>Coordinador: Captura requisitos y diseña los procesos, aplicaciones, bases de datos y otras tecnologías.</p> <p>Títulos:</p> <ul style="list-style-type: none"> - Analista o sistematizador - Experto del conocimiento - Desarrollador - Administrador de la BD 	<p>Fase: Planificar</p> <p>Construye los procesos de calidad dentro del diseño de la base de datos y planifica los procesos con las mejores prácticas para prever errores, incluyendo mejoras en el diseño y la arquitectura.</p>
<p>Productor: Captura, crea o mantiene los datos como parte de sus funciones.</p> <p>Tipos: Interno: Responsable de la entrada de datos en la base de datos o la aplicación Externo: Origina los hechos de su institución</p> <p>Títulos:</p> <ul style="list-style-type: none"> - Tributantes - Personas de la DEMF con potencial para entrar y rectificar datos 	<p>Fase : Obtener</p> <p>Obtener la fuente original de la información.</p>
<p>Desarrollador: Soporte técnico.</p> <p>Títulos:</p> <ul style="list-style-type: none"> - Desarrolladores - Administrador de la BD - IT soporte 	<p>Fase: Almacenar</p> <p>Tecnología (IT):</p> <ul style="list-style-type: none"> - Hardware, software, redes, seguridad para almacenar y compartir datos. - Códigos y consultas para el acceso y mantenimiento a los datos conforme a las regulaciones.
<p>Experto: Requiere y utiliza los datos como parte de sus funciones o en la ejecución de un proceso.</p> <p>Títulos:</p> <ul style="list-style-type: none"> - Especialistas de la DEMF - Usuarios del sistema (personal de las diferentes direcciones del BCC) 	<p>Fase: Aplicar</p> <p>Quienes obtienen los datos pueden ser personas diferentes a quienes lo utilizan y pueden no estar de acuerdo con los expertos, lo que a menudo resulta un problema de calidad.</p>
<p>Administrador: Responsable de la administración del sistema operacional.</p>	<p>Fase: Disponer</p> <p>Manipulando registros: Eliminar datos o</p>

<p>Títulos:</p> <ul style="list-style-type: none"> - Especialista Admón del sistema - Administrador de la BD 	<p>registros es un proceso natural como parte del mantenimiento. En esta fase también se incluye el archivo y la recuperación de datos.</p>
--	---

Captura de datos y Plan de evaluación

1. Priorizar los datos que serán evaluados y ajustarlos en el alcance del proyecto, esquema y recursos. Criterio de selección y la secuencia para la evaluación.
2. Desarrollar un plan de captura que incluya: el método de acceso a los datos, herramientas requeridas, formato de salida, personal involucrado en las actividades y método de muestra. La siguiente tabla permite desarrollar el plan de captura.

Tabla 2.2. Elementos del Plan de captura de datos

<p>Descripción de la población</p>	<p>Describe los tipos de registros que serán extraídos, definiendo las tablas de la BD para el criterio de selección. Incluir cualquier consideración importante: cuando fue creado o actualizado.</p>
<p>Criterio de selección</p>	<p>Describe como se seleccionó la población, la lista de tablas y nombres de atributos (puede incluir instrucciones de SQL). Considerar la edad de los registros. Considerar en el modelo de datos cómo las relaciones afectan la extracción.</p>
<p>Tipo de acceso o extracción</p>	<p>Utilizar variantes: extraer a ficheros planos, a tablas y organizar en área de trabajo o conectarse directamente a la base de producción y copiar datos a un ambiente de prueba.</p>
<p>Datos adicionales (elementos o tablas)</p>	<p>Identificar elementos adicionales a capturar que no serán probados para la calidad, pero necesarios para:</p> <ul style="list-style-type: none"> • Referencias: descripciones de códigos, tablas asociadas, • Identificación: registros únicos, referencias cruzadas.
<p>Responsabilidades</p>	<p>Responsable de documentar las especificaciones de la captura y del completamiento.</p>
<p>Método</p>	<p>Especificar la herramienta utilizada o desarrollada (asegurarse que es un método válido). Para obtener una muestra tener en cuenta: Tamaño: mínimo números de registros para obtener resultados correctos, Estabilidad: si el incremento de tamaño produce los mismos resultados.</p>

El objetivo de esta etapa es analizar, interpretar los resultados, documentar lo aprendido y observado para hacer las recomendaciones iniciales. Sintetizar los resultados para su evaluación e interpretación con los siguientes aspectos:

1. Preparar para el análisis: Tener respuestas claras a las interrogantes de qué se va a evaluar, cómo, quién y cuándo (qué suposiciones fueron aplicadas).
2. Conducir el análisis: Revisar y conducir los resultados identificando patrones y tendencias, chequear información ausente, omisiones, datos menos o más críticos, resultados esperados o inesperados.
3. Esbozar conclusiones y recomendaciones: Las recomendaciones resultantes de impacto en la organización deben ser consideradas preliminarmente La siguiente plantilla permite registrar los resultados.

Plantilla 2.3. Registro del Impacto del análisis

Impacto del análisis (Resultados)					
No.	Tipo/categoría de Evaluación	Lecciones u Observaciones	Impacto Conocido/Probable	Recomendaciones iniciales	Comentarios

Como mecanismo de evaluación y control para la etapa se puede utilizar una lista de comprobación que responda a las preguntas siguientes:

¿Se ha analizado el ambiente de la información (los datos pertinentes, procesos, las personas y la tecnología) al nivel apropiado de detalle para apoyar las metas del proyecto?

¿El ciclo de vida de la información se ha documentado al nivel de detalle apropiado?

¿Se ha identificado algún problema con los permisos y acceso a los datos?

¿Se ha dirigido una valoración de calidad de datos?

¿Se tienen los resultados del análisis, las lecciones aprendidas y las causas raíces iniciales documentadas?

¿El plan de comunicación se ha puesto al día?

La salida de las actividades de esta etapa, el diseño de la captura de datos y la aproximación al plan de evaluación es a su vez la entrada a las dos subsiguientes etapas.

2.3.3 Etapa 3. Evaluar calidad de los datos

Las dimensiones de calidad de datos son empleadas para definir, medir y gestionar la calidad de los datos y la información. Primero se toma la decisión de cuáles dimensiones serán valoradas logrando un balance entre las necesidades y los recursos disponibles;

segundo, se completa la evaluación para las seleccionadas; y tercero, se sintetizan los resultados buscando las correlaciones entre ellas y priorizando aquellas que aporten más valor teniendo en cuenta los objetivos previstos y que los costos no sean prohibitivos teniendo en cuenta la salida de la fase anterior.

Evaluar múltiples dimensiones realmente se hace menos complejo al seleccionar aquellas que soportan la prioridad. Se analizarán los factores de calidad para algunas dimensiones, pero las métricas no serán objeto de estudio, estas pueden ser útil para demostrar la dirección del procedimiento encontrando los hechos, calificar los datos y mostrar las mejoras (Loshin, David, 2006), pero no identifican o resuelven los problemas, sólo proporcionan un indicador de los problemas que existen (Olson, Jack E., 2003). A continuación se detallan las dimensiones relevantes y de interés.

Especificación de los datos

Esta dimensión se refiere a dirigir los esfuerzos para coleccionar, evaluar especificaciones y reglas, instruye para ingresar los datos, diseñar programas de carga de datos (Calero, C., Caballero, I., Serrano, M. y Piattini, M., 2003; Etcheverry, L., Gatto, P.y Tercia, S., 2005), actualizar la información y desarrollar aplicaciones. La tabla siguiente contiene una lista de ejemplos a considerar cuando se evalúa la dimensión.

Tabla 2.3. Especificaciones de Calidad de los datos.

<i>Especificación</i>	<i>Consideraciones</i>
Datos estándar	<p>Nombres de tablas y atributos:</p> <ul style="list-style-type: none"> - Comparar los nombres de la actual estructura física para nombrar convenciones (tablas, vistas, atributos). - Asegurarse que cualquier abreviatura usada en nombres es aceptada como estándar. <p>Entrada de datos:</p> <ul style="list-style-type: none"> - Reglas cuando los datos son ingresados: minúsculas, mayúsculas, signos de puntuación, alfanuméricos, valores mínimos y máximos.
Modelos de datos	<ul style="list-style-type: none"> - Revisar el modelo de datos para asegurar que entidades y relaciones de datos son consistentes y usados. - Identificar como el modelo de datos está siendo comunicado y usado. - Asegurarse que el nombre de la estructura es consistente con el de la convención.

Reglas del negocio	<ul style="list-style-type: none"> - Revisar las reglas para precisar y completar definiciones como datos ausentes, datos ilegales, identificar duplicados, u otras anomalías. - Buscar declaraciones implícitas o explícitas de cómo y cuándo un registro o atributo es tratado a través del ciclo de vida.
Metadatos	<p>Definiciones de datos:</p> <ul style="list-style-type: none"> - Asegurarse que cada atributo y tabla tiene su definición y que a su vez es completa y precisa - Asegurarse cuando los campos pueden considerarse mandatorios, opcionales o condicionales
Datos Referenciales	<p>Dominio y formato para la lista de valores:</p> <ul style="list-style-type: none"> - Chequear si la lista de valores contiene sólo los válidos y determinar si está completa.

Fundamentos de Integridad de datos

Las restantes dimensiones se desarrollan sobre lo aprendido con los fundamentos y especificaciones al proveer el conocimiento necesario mediante técnicas analíticas que busque pero no cambie los datos para descubrir la estructura, contenido y calidad.

La minería de datos (Rahm & H, 2000, pp. 3-13), se define como el análisis y descubrimiento de conocimiento a partir de datos, ayuda a descubrir patrones en conjuntos grandes de datos. Se utilizan modelos descriptivos, los cuales incluyen clasificación (agrupa los datos de acuerdo a similitudes o clases), sumarización, descubrimiento de reglas de asociación o derivar dependencias funcionales y descubrimiento de modelado de datos. Para Fayyad y otros (1996) "La minería de datos es un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos".

El perfil de datos según Olson (2003) para descubrir la verdadera estructura, contenido y calidad de una colección de datos puede llevarse a cabo por una herramienta comercial existente en el mercado o por otra propia mediante consultas de SQL o algún tipo de programa generador de reportes o de análisis estadístico sobre los datos previamente capturados. Existen dos tipos de perfiles: de formato y de contenido, siendo el primero importante porque elimina las ambigüedades sobre el contenido.

El perfil de formato consiste en el análisis de los datos sin tener en cuenta su significado, el análisis se realiza de forma semi-automática y masiva. El perfil de contenido analiza con profundidad el dato y su significado. Existen varios tipos de perfiles, a nivel de columnas y multicolumnas analizando dependencias y redundancias.



Figura 2.7. Perfil de datos, indicadores de calidad

En muchas ocasiones actividades suplementarias como la búsqueda en profundidad de reglas de la organización y relaciones, requieren de intervención humana y trabajar fuera de las herramientas (Chisholm, M., 2004), siendo más efectivo el uso del conocimiento profundo de los datos como hechos, características o fenómenos observables (Levy, 2004) de la organización, pues ocurre que se gasta más tiempo analizando los resultados del perfil de datos en lugar de escribir algunas consultas. La tabla 2.4 describe varias características de los datos al ser sometidos a pruebas, sus definiciones, ejemplos, análisis y acciones potenciales a tener en cuenta en la integridad de los datos.

Tabla 2.4. Pruebas, análisis y acciones para la integridad de los datos

<i>Pruebas</i>	<i>Análisis y acciones de la integridad de los datos</i>
Completamiento (una medida de la cantidad y porcentaje de atributos que no contienen valor)	Es basado solamente en la existencia de un valor. Un análisis adicional es necesario para determinar si el valor es válido, para interpretar los resultados es necesario conocer cuáles atributos son opcionales, mandatorios o condicionales.

Nulos (una medida de la cantidad y porcentaje de atributos que no contienen nada)	Nulo es lo opuesto de completamiento, el mismo análisis anterior es aplicado buscando lo opuesto del anterior punto de vista.
Validez	Definir qué significa válido para cada atributo y la variación de un atributo a otro. Incluye formatos o patrones, códigos válidos, tipo (alfabético/numérico), reglas del negocio, estándares en la entrada de datos.
Tamaño o Longitud	Buscar las diferencias entre el actual tamaño y el esperado inferido por alguna técnica de perfil de datos.
Rangos de mínimo y máximo	Buscar valores fuera de lo documentado o un rango inesperado. Valorar fechas de aperturas de instrumentos de pagos y sus vencimientos, rangos en tipo de cambios para monedas extranjeras, etc.
Contenido	Comparar columnas o nombres de campos con los datos. ¿El campo contiene el dato esperado? - ejemplo fechas, variaciones.
Precisión	Para valores numéricos determinar el valor correcto del lugar decimal para la precisión deseada.
Consistencia (pruebas para atributos en un mismo registro. Integridad referencial)	Buscar consistencia de los datos dentro de un registro y a través de los registros. Buscar al nivel de reglas del negocio para comprender las relaciones y dependencias de otros atributos.
Reglas del negocio	Determinar si las reglas de datos o del negocio no están embebidas en la estructura y requieren ser forzadas por procedimientos o programas lógicos.

Exactitud

La dimensión de exactitud es base para la calidad de los datos. Si los datos no son correctos, otras dimensiones son menos importantes. Tiene en cuenta si los valores guardados para un atributo son los valores correctos y estos deben ser representados en una forma consecuente e inequívoca teniendo en cuenta una fuente autorizada de referencia para comparar. La técnica más importante para la exactitud es el perfil de datos, Olson (2003) lo define como “el uso de técnicas analíticas que descubren la estructura, volumen y calidad de los datos”, no obstante para otras referencias consultadas, el perfil no identifica inexactitud, puede mostrar la presencia de errores (Taylor, J. & Cihon, C., 2004) pero no la ausencia ni el número de errores, por lo que esta técnica será considerada en la dimensión de integridad.

Sin embargo, la exactitud de los datos maestros y referenciales inevitablemente se deteriora, por tanto, es imprescindible establecer un proceso de identificación con una fuente autorizada de referencia que trascienda los límites de los diferentes procesos del negocio (operativos, de colaboración, de análisis y gestión) de la organización y que a su vez sea accesible.

Consistencia y Sincronización

La consistencia se refiere al hecho que los mismos datos usados en varios lugares deben ser equivalentes, el mismo dato debe representar el mismo hecho. La equivalencia es el grado en que los datos guardados en lugares múltiples (para nuestro caso en múltiples modelos informativos) son conceptualmente iguales, esto indica que los datos tienen igual valor y significado. Para Larry English (1999) la sincronización es “el proceso de hacer los datos equivalentes”.

Consistencia y sincronización pueden utilizar las mismas técnicas usadas en el análisis de la integridad simplemente usándola en conjunto de datos múltiples y comparando los resultados. La dimensión de consistencia alcanza capturar la violación de reglas semánticas definidas para un conjunto de atributos que pueden considerarse datos de tuplas de diferentes tablas relacionadas. Por ejemplo para los modelos informativos de Pasivos Externos: Deudas al cierre del mes (1001) y Flujos - movimientos del mes (1002) los atributos que representan la entidad, la identificación del contrato, el proveedor extranjero con quién se contrae la deuda, la moneda y el país del proveedor tienen que ser equivalentes en ambos modelos.

Sin embargo donde los datos no son relacionales las reglas de restricciones pueden aún ser definidas de forma muy similar y expresadas como condición de error, estas son denominadas reglas editables y pueden resultar menos poderosas al no estar expresadas en el modelo relacional.

Disponibilidad y oportunidad

El valor de los datos cambia en el tiempo y siempre existe una brecha en el momento que los datos son capturados electrónicamente para ser actualizados en una base de datos y hacerlos disponibles para el uso. Para ello es importante conocer los procesos que influyen

y el tiempo estimado que la fuente de datos es validada en el SIB, luego es actualizada y almacenada y si hay algo que pueda ser modificado para mejorarlo y que los datos sean oportunos.

Facilidad de uso y conservación

Teniendo los datos disponibles en el repositorio para que ellos puedan ser utilizados puede no ser igual de fácil, determinada información puede estar disponible en las varias tablas de la bases de datos, pero la habilidad de reunir la información para elaborar un informe no es una cuestión simple. También resulta importante conocer que tan fácil o difícil resulta mantener los datos para la toma de decisiones. La facilidad de uso y conservación o mantenimiento están estrechamente relacionadas con la disponibilidad y la oportunidad, pueden analizarse conjuntamente ambas dimensiones.

2.3.4 Etapa 4. Evaluar impacto

Para establecer el impacto de los problemas de calidad de los datos en la organización se requieren técnicas que utilizan medidas cualitativas y cuantitativas con el propósito de lograr el apoyo de la dirección para las inversiones de calidad y determinar cuáles son apropiadas de acuerdo a los recursos (Benítez, 2005) y la necesidad de correcciones y mejoras de los datos para prevenir los errores futuros. Para esta etapa se requiere como entrada, las salidas de las etapas anteriores: las metas y de los objetivos con el conocimiento o sospechas de los problemas de calidad, los resultados del análisis del entorno con las lecciones y recomendaciones iniciales y los resultados de la evaluación de la calidad de los datos.

Primero hay que familiarizarse con las técnicas para conocer que requieren y cuáles son las más apropiadas a la situación, teniendo en cuenta el tiempo y los recursos disponibles, valorarlas teniendo en consideración para la calidad la eficiencia operacional, el número de personas responsables para la toma de decisiones y los riesgos que pueden influir negativamente en el éxito del proyecto (problemas de presupuesto, de personal, problemas técnicos, de organización y comunicación) para finalmente sintetizar y combinar los resultados de todas.

Anécdotas

Son ejemplos o colecciones de historias (Silverman, 2006), fáciles y con bajo costo para evaluar el impacto de la mala calidad. Se realizan a través de las experiencias de las direcciones del BCC siguiendo interrogantes específicas: cómo información equivocada afecta la toma de decisiones, cómo se llevan a cabo procesos claves, cómo datos maestros y transaccionales erróneos o campos obligatorios incrementan la revisión y corrección de informes. Los resultados de las anécdotas pueden coleccionarse en una plantilla.

Tratamiento y uso

Una forma fácil para especificar cómo se emplea actualmente la información y cuáles son los planes futuros de su uso. Se puede cuantificar cuantas personas utilizan la información, al igual que otros sistemas operacionales en explotación en el banco que la utilizan con qué frecuencia y mediante qué vía se realizan las solicitudes.

Matriz de Beneficio contra Costo

Esta técnica analiza y valora las relaciones entre costo y beneficio de cada cuestionamiento, recomendaciones o mejoras. El beneficio puede significar el impacto positivo si las recomendaciones o mejoras son implementadas y el costo es el relativo esfuerzo para ejecutarlas. La siguiente figura muestra la matriz para priorizar las recomendaciones o mejoras.

		Alto	
Beneficio	1. Alto beneficio/bajo costo	2. Alto beneficio/alto costo	
	Aspectos que tienen que ser direccionados primeros	Importantes pero costosos. Pueden ser direccionados antes si los beneficios pesan más que los costos	
	3. Bajo beneficio/bajo costo	4. Bajo beneficio/alto costo	
	Tercera opción , a menos que los aspectos puedan ser unidos a los de los cuadrantes 1 y 2	Última prioridad o no considerados. Hacer un nuevo chequeo	
		Bajo	Alto
		Costo	

Figura 2.8. Matriz Beneficio contra Costo - evaluación de resultados

2.3.5 Etapa 5. Identificar causas raíz

El análisis de las causas raíz no pierde de vista el análisis y priorización de todas las posibles causas del problema de calidad, asunto o condiciones. Su objetivo primario es encontrar por qué los problemas ocurren y que puede hacerse para prevenirlos, direccionando recomendaciones para que no sucedan de nuevo y trazando planes de acción encaminados a eliminarlos.

Las posibles causas raíces, las lecciones aprendidas y las recomendaciones preliminares de las etapas anteriores (etapas 2, 3 y 4) son las entradas para conducir la exploración siguiendo dos situaciones importantes: la evaluación de las dimensiones de calidad y la identificación y localización de datos incorrectos para identificar las causas antes de corregirlos y para direccionar el impacto de un problema específico identificado.

Existen algunas aproximaciones (Loshin, David, 2001b; McGilvray, Danette, 2008; Dylan, Jones, 2010) para el análisis de las causas:

Eliminar la práctica a corto plazo de limpieza de los datos

La limpieza de datos (López, B. y Pérez, R., 2002, 2007) es una de las actividades que induce altos costos, es típicamente una actividad localizada que entrega una solución de apuro y que a menudo no hace nada para prevenir los problemas. Es prácticamente imposible construir una herramienta para la limpieza de datos universal que pueda ser utilizada en cualquier contexto y circunstancias (English, Larry, 1996).

Incluir la fuente de datos externos en la investigación

Se pueden tomar las medidas disponibles para asegurar datos libres de defectos en la organización pero finalmente si están fluyendo datos desde tercera partes, esto puede proponer un punto débil. Es necesario asegurar una política de dirección eficaz a los tributantes con el servicio de acuerdos que continuamente supervisen e informen los defectos. Obsérvese Anexo 5.

Uso de los cinco "Por qué" para descubrir las múltiples causas

El entrevistador simplemente pregunta de forma reiterada el "Por qué" una condición particular surge de forma reiterativa. No confiar simplemente en ¿por qué?, también es necesario preguntar qué, cuándo, dónde y cómo para dar un análisis más enriquecido.

Utilizar el diagrama de Pareto

El diagrama de Pareto (Fernández, 1999; Vidaña, 2010) es una gráfica de barras que ilustra las causas de los problemas por orden de importancia y frecuencia (porcentaje) de aparición, costo o actuación. Se pueden detectar los problemas que tienen más relevancia mediante la aplicación del principio de la regla 80/20 (muchos triviales y pocos vitales) que dice que hay muchos problemas sin importancia frente a solo unos graves. Los problemas más frecuentes no pueden tener los impactos mayores en términos de calidad, costo o tiempo. Se recomienda construir dos diagramas, uno teniendo en cuenta (defectos y frecuencia) y otro que valore el impacto (costo y tiempo) para identificar los que ocurren más frecuentemente y tienen el mayor impacto canalizando los esfuerzos hacia los 'pocos vitales'.

Utilizar el diagrama Causa- Efecto

Esta herramienta gráfica conocida también como diagrama de Ishikawa (Fernández, 1999; Vidaña, 2010) o de espina de pescado (*fishbone*), es útil en la identificación de las posibles causas de un problema y representa las relaciones entre algunos efectos y sus causas. Es una de las herramientas más esenciales para el análisis de las causas raíces y se despliega típicamente en una tormenta de ideas. Para determinar las causas, Loshin (2001b) sugiere como buena práctica revisar los problemas crónicos, aquellos que han surgido alrededor de mucho tiempo y fueron ignorados y los problemas agudos que han surgidos recientemente y están colocando nuevas presiones en el actual sistema operacional (pp. 389-391).

2.3.6 Etapa 6. Plan de mejoras

Con el tiempo debido a la evolución de los procesos en el negocio, los expertos e informáticos han desarrollado el control de los errores pero sin solucionar los problemas específicos. Se pueden haber realizado arreglos rápidos para ocuparse de datos defectuosos y cerca de resolver las causas raíces del problema se ha ocasionado un serio impacto. Esta etapa es un punto crítico en el procedimiento, la meta es entender qué actividades y datos manejan los procesos que lleven a los resultados más eficaces para la mejora de calidad de la información y los flujos de datos (Abate, M. & Diegert, K., 1998). La comunicación es esencial para asegurar que las recomendaciones de calidad y la

valoración del impacto, ambas resultantes del análisis de las causas raíces, sean implementadas.

La mejora puede ser implementada por el propio equipo de trabajo debido a los conocimientos y habilidades acumuladas si se busca ganar en rapidez y si las actividades en cortos términos brindan los beneficios, o por el contrario, si el plan requiere de un alto nivel puede ser desarrollado por otros incrementado la comunicación para encontrar similitudes a través de las recomendaciones. De cualquier manera, se debe asegurar que el plan de mejora incluya la corrección, la comunicación, la prevención e identificar el personal que más responsabilidad y sentido de pertenencia tiene en cada tarea mediante la supervisión continua.

2.3.7 Etapa 7. Plan de comunicación

La comunicación es esencial para revelar el éxito del procedimiento en la mejora de la calidad de los datos. Los resultados y acciones deben ser notificados a través del proyecto y en todas las etapas que lo componen. Tiene como propósito educar y elevar el conocimiento de la calidad de los datos; así como la importancia del impacto en el BCC, obtener la ayuda y el soporte necesario de los interesados demostrándole el valor de la información y su mejora, proporcionar visibilidad y asegurar apoyo a los analistas que dependen de la información para el cumplimiento de sus funciones y que estos conozcan cómo pueden guiar la mejora continuamente, ganar la cooperación en la corrección y prevención de errores en los datos en aquellas personas que tiene responsabilidad en los procesos y toman decisiones.

Como instrumento para facilitar la comunicación de las acciones y resultados necesarios del procedimiento, se define una plantilla para uso de todos los miembros del equipo. El llenado se iniciará en el momento que se tienen las lecciones aprendidas y recomendaciones en cada etapa. La información será completada en la medida en que avance el proceso y vayan siendo precisadas y refinadas las acciones. Esta plantilla recogerá:

Audiencia: las personas individuales, equipo de trabajo, organizaciones o departamentos del BCC que necesitan escuchar y serán afectados.

Mensaje y acción: qué se necesita conocer, qué está cambiando, qué acciones hay que ejecutar y cómo la audiencia es impactada.

Medio de comunicación: cuál es el método para comunicar (presentaciones, seminarios, reuniones, correos electrónicos).

Desarrollo y entrega: responsable para fomentar, crear y entregar la comunicación, persona que provee el contenido.

Preparación de la acción: acciones necesarias para preparar y completar la comunicación.

Fecha planificada: fecha proyectada.

Fecha completada: fecha en que la información es completada

2.4 Conclusiones parciales

- El procedimiento elaborado para su aplicación en el SIB tiene características proactivas, porque se prevé el ambiente o entorno y los cambios de la cultura organizacional desde las etapas más tempranas del ciclo de vida para llevar a cabo los objetivos propuestos.
- Como procedimiento para gestión de la calidad de datos, evalúa componentes y otros factores que si no son direccionados ponen en riesgo la calidad de la información: requerimientos y restricciones, responsabilidades y gobernación, mejoras y prevención, la estructura y significando que proporcionan el contexto para los datos financieros, la comunicación y la gestión de cambio asociada al impacto en la organización.
- Su concepción y diseño está basado en las “mejoras practicas” y pautas en obediencia de las características y nivel de calidad de datos requerido por la DEMF.
- La naturaleza iterativa e incremental garantiza el mejoramiento gradual del conocimiento del equipo de trabajo y la definición de un conjunto de actividades para la mejora, minimizando las causas que producen datos defectuosos de forma más objetiva y realista.

- El procedimiento está sustentado en un enfoque holístico y sus cambios a lo largo de todo el ciclo de vida del proyecto es independiente de la metodología de desarrollo adoptada.

CAPÍTULO III. EVALUACIÓN DEL PROCEDIMIENTO EN EL ACTUAL SIB

3.1 Evaluación del Procedimiento en el SIB

El procedimiento elaborado en el capítulo anterior requiere, por un lado, de su aplicación práctica en el actual Sistema Informativo Bancario y de la validación de la aplicabilidad de las "herramientas" y "mejores prácticas" propuestas a partir sus características previamente definidas.

La decisión de aplicar el procedimiento al SIB se fundamentó en la necesidad de desarrollar mejoras a los procesos del actual sistema, en particular al proceso de validación de los datos desde la fuente original, al flujo de las actividades y a su vez evaluar el estado actual de los datos en el repositorio, que cuenta con información histórica que data del año 2000 aproximadamente. El sistema funcionaba en la DEMF desde su concepción, posteriormente se decidió automatizar el módulo manual de validación y se le han solicitados continuos cambios para adicionar y mejorar sus funcionalidades que, lejos de ayudar han ocasionado más cargas y consumo de tiempo de verificación.

3.1.1 Etapa 1. Definir necesidades y especificaciones

Esta etapa se desarrolló según establece el procedimiento para la obtención de la información acerca del dominio del problema y del sistema actual, con el conocimiento práctico real que poseen los miembros de la DEMF y el personal informático involucrado.

Se creó un equipo de trabajo conformado por siete miembros: director de la DEMF (funcionario de alto nivel), dos analistas de la DEMF (expertos del negocio), dos especialistas informáticos (analista y desarrollador) y dos usuarios finales (departamentos de Supervisión Bancaria y Política Monetaria). Como jefe del proyecto se designó a la analista que desarrolló modificaciones y mejoras a la primera versión del sistema operacional, cuyos conocimientos especializados del dominio del problema y características personales para la comunicación, son suficientes para llevar adelante el proyecto y aplicar este procedimiento.

Se evidenció que la composición del equipo propició que los miembros se sintieran responsables por los resultados e hicieran suya la tarea con entusiasmo, las relaciones interpersonales fluyeron adecuadamente a partir del compromiso colectivo con la obtención de un procedimiento de calidad que resolviera sus problemas y que pudiera ser generalizado con éxito. La utilización de entrevistas y la técnica de calidad tormenta de ideas permitió la creatividad con un número grande de proposiciones, en que todos los miembros del equipo se involucraron con sentido de pertenencia en las decisiones; también propició la detección rápida de lagunas informativas, que fue posible superar con la construcción de otras ideas, aclarándolas y combinándolas.

A continuación se detallan las necesidades identificadas y ampliadas en la plantilla que sigue a continuación, con relación a las cuatro problemáticas consideradas:

P1.Fuente externa de datos segura:

- Identificar el origen de la información (entidad tributante) para poder tomar acciones correctivas.

P2.Costos innecesarios o excesivos:

- Detectar desvíos de la información registrada (códigos erróneos, fechas de entregas incorrectas, validar formatos) con respecto a la información real. Estandarizar e intentar corregir datos permisibles de los tributantes a nivel general, para evitar costos innecesarios una vez incorporados los datos al repositorio.

P3.Incrementos de riesgos por datos expuestos sin previo análisis de errores, imprecisiones o inconsistencias:

- Adaptación a nuevos requisitos controlando que todos los datos cumplan con las reglas de validación existentes. Cambios por nuevas reglas de negocio, que exigen la incorporación de nuevos datos y reglas de validación de la información.
- Trasladar los datos al formato estándar actual.
- Detectar y completar la omisión de datos obligatorios y con valores permitidos.

P4.Datos y procesos compartidos:

- Contar con información actualizada y completa que sustenta a otros procesos y sistemas operacionales.

Aproximación del Problema						
No	Problema	Datos	Procesos	Personal y Organización	Herramienta y/o Tecnología	Comentarios
P1	Fuente externa de datos	Modelos que informan los tributantes	Obtener modelos vía correo electrónico	Tributantes: OACE, SBFN y DEMF	Pendiente de definir e implementar	-Chequear dominios y cuentas de correos válidas - Verificar formato válido del adjunto - Copia física del modelo
P2	Costos innecesarios o excesivos	Modelos que informan los tributantes	Validación automática y manual	DEMF e Informáticos	Sistema operacional	- Verificar código del tributante y fechas de entregas del modelo - Validar estructura y orden de campos en el modelo - Estandarizar atributos
P3	Incrementos de riesgos por datos incorrectos	Modelos que informan los tributantes y análisis del repositorio	Validación automática y manual, herramientas analíticas y de transformación de datos	DEMF e Informáticos	Pendiente definir e implementar herramienta y transformación de datos	- Desarrollar reglas, procesos, estándares - Transformar datos
P4	Datos y procesos compartidos	Datos del repositorio	Almacenar, consultar (transformación de datos) y obtener Información de los modelos	DEMF e Informáticos	Sistema operacional, pendiente definir e implementar transformación de datos	- Datos almacenados, actualizados y compartidos - Triggers para mantenimiento - Transformar datos

Se realizó el análisis de factibilidad y sus resultados arrojaron que el proyecto no sólo era viable económica, legalmente y técnicamente, sino que el grado de necesidad era superior al valorado inicialmente. Se comprobó que al menos cuatro direcciones más del BCC (Deuda Externa, Sistemas de Pagos, Supervisión Bancaria y Política Monetaria) tenían las mismas necesidades y estaban interesadas en la gestión de la calidad de la información para el cumplimiento de sus tareas. Los interesados aportaron dos nuevas necesidades a la tercera problemática aún más complicadas: la comprobación del Balance General (modelo 0920) y la conciliación de la información entre modelos afines, este último se decidió realizar solamente un análisis preliminar y posponerlo para la potencial incorporación de una futura etapa de “Implementación de controles” dada su complejidad.

3.1.2 Etapa 2. Analizar entorno de la información

En esta etapa se fue más preciso en el análisis de los datos y sus especificaciones comenzando desde un alto nivel para caracterizarlos acorde a su significado y roles en:

cualitativos o categóricos (almacenadas transaccionalmente en los maestros de Espacio y Concepto), cuantitativos o numéricos (en el maestro de Valor) y de fechas o plazos (en el maestro de Tiempo) hasta obtener la plantilla detallada con las características básicas: reglas, datos estándares, datos referenciales o metadatos, información estructural u otra información significativa.

En el sistema operacional actual (Anexo 2) se cuenta con información acerca de datos referenciales (denominados clasificadores según el lenguaje utilizado en los actuales manuales de usuarios y de explotación del sistema) y que son de utilidad para diversos sistemas implementados en el banco pues contienen lista de valores, descripciones y códigos; datos transaccionales referidos como maestros; datos maestros que describen la metodología de los modelos informativos de manera estandarizada e incluyen asociación a datos referenciales; metadatos que describen la tecnología y estructura de los datos y por último datos temporales sólo para propósitos técnicos que no pueden ser visualizados.

La tecnología y los procesos fueron analizados de conjunto por su interrelación y a su vez mediante el ciclo de vida de los datos. Existen procesos para **Planificar** las actividades relacionadas con los módulos de validación manual y automática del sistema, para **Obtener** los modelos vía correo electrónico de los contribuyentes, para **Almacenar** la información a verificar en tablas temporales a través de procedimientos almacenados que luego será actualizada y archivada en el repositorio mediante el módulo de actualización, para **Aplicar** y obtener información mediante el módulo de consulta aplicando transformaciones a los datos para que puedan ser examinados y para **Disponer** o mantener la información a través del módulo de administración que permite archivar y/o eliminar registros transaccionales, definir nuevas estructuras para los modelos, mantener y actualizar datos referenciales, definir y adicionar o eliminar reglas de verificación caducadas por la evolución e intereses del negocio.

En la caracterización de las personas/organizaciones, roles y responsabilidades se evidenció la falta de habilidades y práctica para lidiar con proyectos relacionados con la calidad de la información. La inexperiencia y falta de preparación del equipo de trabajo es una cuestión a superar para identificar, evaluar y planificar la Gestión de la Calidad de la Información como una herramienta de gestión general que se utiliza para diseñar la calidad en los procesos a prueba de errores y garantizar que estos cumplan los requisitos

esperados, a diferencia de la Gestión de la Información, una disciplina distinta asociada a la gestión empresarial.

Comprender el ciclo de vida de la información y su relación con el entorno permitió las consideraciones adecuadas para la elaboración del plan de captura y evaluación. Los datos contenidos en los maestros (datos transaccionales) son el objeto del plan, valorar los mismos hasta cinco o seis años atrás ha garantizado elaborar criterios sólidos, con seguridad hay segmentos de información en la base de datos muy antiguos que no han transitado por los procedimientos de control vigentes, por lo que no cumplen con ciertas normas.

Se diseñó una herramienta propia a la medida, para la captura de los datos de la fuente operacional y convertirlos a los formatos de utilización mediante consultas y la sentencia “*Pivot*” de SQL que permite la integración y normalización, valiéndose de la información de los metadatos (P_REPDATOS), el maestro que contiene la estructura de los modelos informativos (P_MODELOS) y teniendo en cuenta que las relaciones del actual esquema de datos afectan la extracción, para posteriormente cargarlos en tablas de trabajo, limpiarlos y continuar el análisis. El programa se desarrolló en *MS Visual Studio* con el empleo de *MS.NET Framework Common Language Runtime (CLR)* y el tiempo para su diseño e implementación consumió seis semanas. De otra parte, las herramientas comerciales que ejecutan depuración a los datos, en general, no realizan autónoma y automáticamente este trabajo, sino que requieren la intervención del usuario para elegir la técnica a ser aplicada a los datos, tarea que demanda altos conocimientos técnicos.

Considerar a continuación dos consultas (extraídas de la propia herramienta implementada) como ejemplo para la creación de la tabla y carga de los datos para el modelo informativo 0920 - Balance General:

```
CREATE TABLE dbo.M_0920 (FEC_MODELO datetime constraint fec_not_null not null,
COD_INSTITUCION varchar(5) constraint inst_not_null not null, COD_INDICA varchar(4),
NOM_INDICA varchar(100), COD_BANCO varchar(6), IMPORTECUC numeric(18,2),
IMPORTEUSD numeric(18,2), IMP_MDANAC numeric(18,2)); Create clustered index
Idx_Fec_Inst ON dbo.M_0920 (FEC_MODELO, COD_INSTITUCION)
```

```

INSERT INTO dbo.m_0920 (FEC_MODELO, COD_INSTITUCION, COD_INDICA,
NOM_INDICA, COD_BANCO, IMPORTECUC, IMPORTEUSD, IMP_MDANAC)
SELECT fec_modelo, cod_institucion, cast([00157] as varchar(4)) as [00157], cast([00005]
as varchar(100)) as [00005], cast([00105] as varchar(6)) as [00105], cast([00190] as
numeric(18, 2)) as [00190], cast([00150] as numeric(18, 2)) as [00150], cast([00151] as
numeric(18, 2)) as [00151] FROM
(SELECT fec_modelo, cod_institucion, fila, cod_dato, valor from (select fec_modelo,
cod_institucion, fila, cod_dato, valor FROM dbo.m_concepto where cod_modelo = '0920'
UNION SELECT fec_modelo, cod_institucion, fila, cod_dato, valor FROM dbo.m_espacio
where cod_modelo = '0920' UNION SELECT fec_modelo, cod_institucion, fila, cod_dato,
valor FROM dbo.m_tiempo where cod_modelo = '0920' UNION SELECT fec_modelo,
cod_institucion, fila, cod_dato, valor FROM dbo.m_valor where cod_modelo = '0920') AS
a) p pivot (max(valor) FOR cod_dato in ([00157], [00005], [00105], [00190], [00150],
[00151])) as pvt ORDER BY fec_modelo, cod_institucion, fila

```

Durante la exploración y valoración para la limpieza de datos en el repositorio del SIB (corregir, estandarizar y completar los datos) se identifican valores perdidos para completar campos vacíos con datos referenciales, se identifican valores ilegales examinando la cardinalidad y los valores máximos o mínimos fuera del rango permitido, códigos incorrectos que pueden ser re-codificados y la posibilidad de modificar el contenido de algunos campos con errores típicos (diferentes códigos con el mismo significado). Se llega a la conclusión que el costo asociado a la limpieza de datos no agrega valor, simplemente lleva los datos a un estado de utilidad y de relativa confianza.

Como resultado de la evaluación se puede señalar, que no es posible detectar todos los errores en cuanto a rangos de fechas permitidas (vencimientos, términos, apertura de instrumentos de pagos), estas deben ser valoradas de acuerdo a la metodología de los modelos informativos. Los datos de tipo numérico, específicamente los saldos, no reportan elementos que permitan detectar errores, las reglas del negocio no establecen ninguna limitante para esta cantidad, al no ser algunas reglas de cómputos o de inferencia del proceso de validación a través de una o más columnas de una tabla. La detección de duplicados está diseñada generalmente (según la literatura) para ser aplicada a cadenas de texto y estas actualmente son reemplazadas por los datos referenciales (nombre de instituciones, organismos, etc.) durante el proceso de validación para evitar

inconsistencias, se garantiza no contener registros duplicados que corresponden a la misma entidad tributante para una misma fecha mediante el proceso de actualización. Debe tenerse presente que no todo valor nulo es un faltante pues de acuerdo con la naturaleza del dato, es posible que éste no sea obligatorio a partir de un momento determinado y por tanto no sea un error, esta labor se puede realizar utilizando herramientas de perfil de datos. Más importante aún, se concluye que la mayoría de los problemas de calidad no son debido a la inadecuada dirección de los procesos existentes, sino al esquema (definición de la estructura de la base de datos) inadecuado, contextualmente incompleto o inexacto y donde la integridad de los datos almacenados se pierde (integridad de entidad y referencial). Se detectan reglas del negocio embebidas en los códigos fuentes de los programas, lo que dificulta el mantenimiento y no pueden ser gestionadas por el personal del negocio.

Impacto del Análisis(Resultados)					
No.	Tipo/categoría de Evaluación	Lecciones u Observaciones	Impacto Conocido/Probable	Recomendaciones iniciales	Comentarios
1	Funcionalidad del dato	Inconsistencia, imprecisión del dato	Error Grave	Establecer reglas del dominio	
2	Impacto Económico	Informaciones incorrectas a organismos de alto nivel	Error Grave	Limpieza y depuración de datos	
3	Toma de Decisiones	Análisis e informes incorrectos	Error Grave y Menos Grave	Identificar inexactitud de los datos	Gasto de tiempo verificando los datos
4	Dependencia de otros datos	Informaciones entre modelos afines inconsistentes	Error de Gravedad Media	Establecer dependencias funcionales y reglas de integridad y consistencia	
5	Reclamaciones	Se exige entrega de información adicional a las organizaciones	Error Menos Grave	Toda entrega debe ser enviada a través del sistema operacional	No quedan registradas en el repositorio. Sin metodología asociada

3.1.3 Etapa 3. Evaluar calidad de los datos

A partir de las conclusiones y recomendaciones resultantes del análisis del impacto en la organización en la etapa anterior, se determinan las dimensiones vinculadas al incumplimiento de las características de calidad que deben poseer los valores de los datos, ya sea a través de la existencia de registros con esta información, por experiencia de situaciones que se han presentado en concordancia con los criterios dados por especialistas o por la reflexión de otras situaciones que potencialmente pueden surgir El

resultado de este proceso es la lista de dimensiones que se utilizará para evaluar la calidad de los datos según el grado de importancia.

Especificación de los datos

Iniciativas proactivas para el manejo de la calidad de los datos comienzan con verificaciones en la entrada de datos en el actual módulo de validación, son la primera línea de defensa contra datos erróneos: aseguran que todos los campos posibles se comprueben contra datos referenciales, comparan los nombres de atributos contra lo aceptado como estándar, se aplican reglas de datos para verificar validez y se comprueba que la estructura del modelo informativo es consistente con el de la convención. Sin embargo, no se asegura en los metadatos cuando los campos pueden considerarse mandatorios u opcionales de acuerdo a condiciones documentadas y no chequean rangos de valores autorizados para aquellos datos donde sea permisible. Se detectan campos que no se corresponden con la longitud esperada en los metadatos (campos que presumiblemente no fueron sujetos a las reglas del negocio en su momento) y valores nulos. Para el modelo informativo 0301 los atributos: código de la institución representan correctamente el 98.32% y el código del organismo el 99.14%. Se detecta para el atributo origen de los recursos un 15% de nulos.

Integridad de datos

Muchas reglas de integridad no están definidas en la base de datos. Se puede decir que es un problema que no está resuelto a nivel de la base de datos y al ser necesarias se controlan a nivel de procedimientos o funciones que se aplican a los atributos durante la validación. Para su estudio existen tres factores:

Integridad de dominio: controla reglas sobre el contenido de un atributo de una tabla. Por ejemplo para el modelo 0915, la fecha de vencimiento de la letra de cambio tiene una duración de 3 años y 6 meses que no es controlada. Para el modelo 1001, la fecha de emisión del instrumento de pago (letras, pagarés, cartas de créditos) contiene fechas futuras a la instituida y el término de vencimiento de la deuda (expresado en fracciones de días, meses, años) muestra en ocasiones valores negativos.

Integridad intra-relación: controla reglas sobre la relación entre varios atributos de una misma tabla. Por ejemplo para el modelo 0101, el total de cuentas por cobrar vencidas

debe ser menor que el total de cuentas por cobrar en MLC (Moneda Librementemente Convertible) y MN (Moneda Nacional) y a su vez mayor que cero.

Integridad inter-relación: controla la satisfacción de reglas entre atributos de diferentes tablas. La regla más común es la llave foránea, la cual controla que el valor de una celda esté incluido en la llave primaria de la tabla referenciada. Este factor sólo es mostrado entre dos tablas del modelo: el repertorio de datos (P_REPDATOS) y maestro que describe la estructura de los datos referenciales o clasificadores (P_CLASIFICADOR).

El perfil de formato no fue objeto de análisis en esta dimensión, ha sido completado en la evaluación de la dimensión anterior. Se realiza el perfil de dependencia funcional como método inductivo para el análisis de las reglas de integridad y examinar llaves primarias. Para los modelos 1001 y 1002 el atributo prestamista depende de un 95.99% del conjunto de atributos: código de la institución, código del país, número del contrato y sigla de la moneda.

Exactitud

La dimensión requiere comparar los datos con una fuente de referencia autorizada, su conducción es usualmente un proceso manual demorado y susceptible a errores, característica que impulsa la correcta decisión de los elementos a evaluar. Es la dimensión más dramática y complicada que no se puede conseguir perfeccionar en un cien por ciento, pero en aquellos elementos donde no es práctico y resulta costoso por el tiempo disponible, se puede mejorar mediante una estrategia continua con el empleo de técnicas analíticas. Se definen los siguientes factores:

Corrección Semántica: indica qué tan bien se representan los datos en relación al dominio: i) datos que no se corresponden con ningún valor real; ii) datos que corresponden a un valor equivocado. Por ejemplo, la información registrada en la base de datos de un organismo o institución inexistente, que fue dado de baja por la ONE (Oficina Nacional de Estadísticas) a partir de una fecha o que fue fusionado, y por tanto se re-codifica.

Corrección Sintáctica: indica si los datos están libres de errores sintácticos si satisfacen las reglas y restricciones impuestas: i) errores de valores que pueden estar fuera de rango para las fechas, se reportan años anteriores al 1900; ii) valores embebidos, en un mismo atributo que corresponden a múltiples atributos.

Precisión: indica que tan detallados están los datos. Calcula la diferencia que hay entre el nivel de detalle del dato registrado y el esperado. Por ejemplo, la cantidad de decimales que se reconoce en una tasa de interés o el monto asociado a un tipo de moneda específica.

Los datos evaluados en esta dimensión de manera manual fueron los datos referenciales denominados clasificadores o nomencladores. Su análisis arrojó un 10% de errores de acuerdo al nivel de tolerancia, fue la dimensión más tangible de tratar y fácil de mejorar al no requerir reingeniería de procesos ni reestructuración de la organización. Sin embargo el análisis del nomenclador (C_INSTITUCION) requirió más tiempo y esfuerzo al contener 17, 840 artículos aproximadamente (se conservan códigos para series históricas).

Los metadatos almacenados (P_REPDATOS) mostraron códigos redundantes para una misma interpretación, incorrectos e inconsistentes para asegurar la calidad de los datos y muy pocas reglas de integridad declaradas y almacenadas. Por ello fue importante analizar instancias actuales para obtener metadatos reales (reingeniería). A continuación algunos ejemplos descubiertos mediante consultas y perfil de datos:

- Para representar el código de un país aparecen dos códigos (00241, 00311).
- Para especificar totalidad de los cobros realizados se duplican los códigos (00051,00197).
- Códigos que no se utilizan en los datos maestros (P_MODELOS y P_CLASIFICADOR)
 - 00121 - Financiamiento Aprobado
 - 00061 - Mes
 - 00062 - Año
 - 00063 - Saldo_0_30_MLC_Cuba
 - 00064 - Saldo_0_30_MLC_Exterior
 - 00065 - Saldo_0_30_MN_Cuba
 - 00066 - Saldo_31_60_MLC_Cuba
 - 00067 - Saldo_31_60_MLC_Cuba
 - 00068 - Saldo_31_60_MLC_Exterior
 - 00069 - Saldo_31_60_MN_Cuba

- Código 00033 - Observaciones, no se utiliza en los maestros anteriores y almacenamiento asociado a una tabla transaccional inexistente.
- Código 00314 - FEC_E_CC no se especifica en cuál tabla transaccional es almacenado. No tiene descripción asociada.

La mayor cantidad de campos son almacenados en el maestro de valor para un 63.17%, seguido del maestro de concepto con un 23.80%. Un total de 281 códigos son almacenados en los maestros transaccionales, representando el 89.2% del total de 315 códigos, el resto conforman campos de los nomencladores.

Consistencia y Sincronización

Esta dimensión será prudente evaluarla para el objetivo que se persigue, analizando consistencia a través de tablas (o modelos informativos afines), una de las necesidades que emergió en la primera etapa y desestimada a partir de la información recopilada que arrojó su complejidad. En la dimensión de integridad también se valoró la consistencia pero para atributos en un mismo registro o entre varios registros, así como el análisis de dependencias entre atributos a través de reglas de datos o del negocio.

En el próximo paso con el análisis de las técnicas de impacto para la organización y como resultado de evaluar su costo – beneficio, se decidirá su adopción para realizar perfil buscando redundancias y problemas de integridad referencial entre tablas relacionadas. El análisis para los modelos informativos 1001 y 1002 revela que varias columnas comparten el mismo juego de valores, en la tabla 1002 no es llave foránea y en la tabla 1001 puede especificarse como llave primaria.

Disponibilidad y oportunidad

El valor de los datos cambia en el tiempo, existe un momento en que son capturados electrónicamente o manualmente, verificados y almacenados temporalmente para luego ser actualizados en el repositorio y hacerlos disponibles para el uso.

Para ello fue importante examinar los procesos que influyen y el tiempo estimado de sus actividades con el objetivo de mejorarlos para disponer de datos oportunos. El proceso de captura y copia física de un modelo informativo apenas demora de 20 a 30 segundos, la verificación está en dependencia de la cantidad de atributos, cantidad de registros y las reglas de datos que tienen que ser validadas a través de una o más columnas del modelo. Por ejemplo, el tiempo estimado para el modelo 2400 (16 segundos), para el modelo 1002 (2 minutos y 14 segundos) y el modelo 0400 (1 hora) aproximadamente para un gran número de registros.

El proceso asociado con el módulo de actualización de los datos desde las tablas temporales a las transaccionales se ejecutaba cada tres horas aproximadamente de manera automática, con la posibilidad de forzarlo manualmente por el/los administrador(es) del sistema. No obstante fue necesario disminuir estos tiempos a menos de una hora, para ello se mejoró el proceso introduciendo cambios mediante el control de banderas (*flags*) que garantizarán que los procesos de validación y actualización no colapsen, dando prioridad a la actualización.

Otro aspecto importante durante el análisis de esta dimensión resultó la frecuencia de actualización de unos de los datos referenciales más importantes mencionado anteriormente (C_INSTITUCION) que requiere de actualización mensual oportuna para evitar rechazos de modelos informativos ante entidades no actualizadas o que causaron baja para la ONE.

Facilidad de uso y conservación

Los datos almacenados en el repositorio están disponibles en varias tablas y necesitan ser transformados a su formato y estructura de origen (todos están depositados en el mismo formato de cadenas de caracteres de longitud fija en las tablas transaccionales de acuerdo al indicador económico) para que puedan ser utilizados, por tal motivo la recuperación es a su vez un problema en cuanto a tiempo de espera y volumen. El actual módulo de consulta, de utilidad continua para el usuario, no permite su mejora y está sujeto a constantes quejas y reclamaciones en particular cuando se pretende consultar series históricas.

Trazabilidad

Aparece esta dimensión por el interés creciente, tanto a nivel de la DEMF como de la institución de conocer información adicional de los datos. Para el sistema que almacena datos financieros, es de mucha utilidad identificar cuando se han realizado cambios en los datos, quién los realizó y de qué fuente proviene la información, al derivarse otras propiedades, tal como la confiabilidad de su fuente. El factor analizado es la verificabilidad de la información: que representa cuán localizables son los datos de actualización. Por ejemplo en sistemas de auditoría, la verificabilidad indica el grado en el que se puede identificar la fuente de actualización de los datos y la historia de las modificaciones.

En el actual sistema fue objeto de análisis la tabla (M_ENTSAL) que muestra esta información y refleja el estado del modelo (validado, rechazado, no reportado y errores de otro tipo), a lo que se sugiere agregar un atributo que contenga una descripción global de los problemas detectados en caso que proceda.

De manera general, respecto al análisis de las dimensiones se obtuvo como resultado que todos los errores de integridad fueron graves o de gravedad media, en el caso de la exactitud el 10% de los errores fueron graves, los errores de consistencia todos resultaron de gravedad media; parece suceder que los problemas de integridad producen las consecuencias más graves que provocan datos incorrectos, incompletos o perdidos y relaciones mal resueltas, siendo también los más frecuentes. El resto de las dimensiones se enfocan en cómo el grupo de usuarios interpreta y utiliza los datos.

3.1.4 Etapa 4. Evaluar impacto

Para establecer el impacto en la organización de los problemas de calidad de los datos se requieren medidas cualitativas y cuantitativas. Mientras mayor sea el nivel de calidad de los datos requeridos se demanda un elevado costo y tiempo de esfuerzo considerable.

Anécdotas

Los resultados de tres anécdotas significativas fueron coleccionados luego de entrevistar al personal de cinco departamentos del BCC. Se muestra la información resultante de una de las anécdotas en la siguiente plantilla:

Título: Especificaciones incorrectas en los datos maestros	
Dato: Registro del Modelo informativo 0920 (Balance General Supervisión a distancia)	Persona: Directora del departamento de Supervisión Bancaria
Procesos: validación y verificación	Tecnología: sistema operacional
Escenario: nombre y orden del dato incorrecto	
Impacto (cuantificar si es posible): <ul style="list-style-type: none"> • Provoca la evaluación de regla de inferencia y de cálculo del dato errada • Imposible comprobar si el dato es aceptable (validez y exactitud) • Se rechaza el modelo (desconfianza de la información de origen) Recomendaciones iniciales: Correcta administración en el sistema operacional del maestro (P_MODELOS)	
Ofrecida por: Idania Torre	Fecha: 03/07/2011

Tratamiento y uso

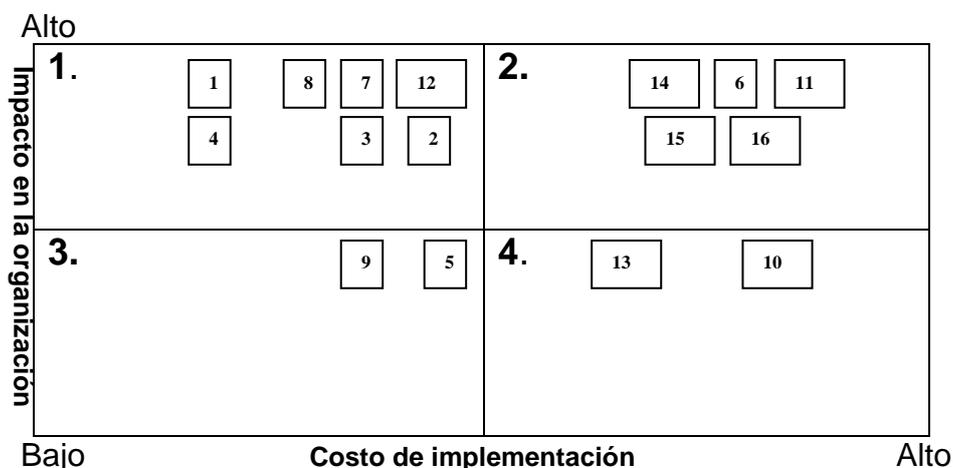
Actualmente los datos almacenados son utilizados en la mayoría de las direcciones del BCC con el propósito de examinar modelos informativos, confeccionar informes a los altos organismos y para la toma de decisiones en la ejecución de acciones, teniendo en cuenta nuevas políticas del sector bancario. Se exponen algunas derivaciones relacionados con el tratamiento y su uso:

- Balanza de Pagos: Es un estado estadístico de las transacciones económicas de la economía y el resto del mundo. Se refieren a bienes, servicios y renta.
- Informe del Financiamiento del SBFN a la economía interna: Financiamientos vencidos, renegociados y saldos pendientes de amortizar en MLC y CUC.
- Ingresos y egresos de cuentas corrientes de personas naturales y jurídicas.
- Circulación monetaria: estudio de la base monetaria.
- Central de riesgo: Cheques deficientes, letras de cambio no pagadas o pagadas parcialmente, transacciones rechazadas y medidas aplicadas a infractores.

La DEMF tiene como visión elaborar un Boletín Estadístico que contenga informes y tablas con información consolidada y enriquecida, además de futuros análisis de series históricas para trabajos de econometría y otros trabajos adicionales macroeconómicos.

Matriz de Beneficio contra Costo

En esta etapa fueron especificadas dieciséis recomendaciones y/o mejoras al valorar las relaciones entre costo y beneficio de cada acción. Las recomendaciones o mejoras se revelan en la matriz de beneficio contra costo.



Etapa 1:

1. Conocer fuente origen de los datos.
2. Corregir errores de la fuente original.
3. Estandarizar datos.
4. Detectar y completar omisión de datos.
5. Cambios en reglas de datos e implementación de nuevas reglas del negocio.
6. Transformar, actualizar y asegurar los datos para su uso.

Etapa 2:

7. Mejorar procesos y flujos actuales.
8. Entrenamiento y capacitación del personal involucrado.
9. Estudio de herramientas comerciales e integradas para la limpieza.
10. Emplear métodos y herramientas de limpieza.
11. Nuevo diseño del esquema lógico de datos.
12. Perfeccionar el actual sistema.
13. Infraestructura desacoplada del código fuente para las reglas del negocio.

Etapa 3:

14. Seleccionar y priorizar las dimensiones de calidad a evaluar.
15. Perfilar datos (atributos y registros de una misma tabla)
16. Perfilar datos (para modelos afines entre tablas)

Conjuntamente a las recomendaciones, se discuten y ordenan jerárquicamente las dimensiones de calidad de datos para su priorización y análisis. La dimensión Consistencia y Sincronización tiene un costo de implementación significativo sobre todo por el tiempo disponible y el esfuerzo requerido por los analistas de datos para interpretar los resultados, por tanto se decide evaluar sólo para los modelos más significativos que influyen en la toma de decisiones.

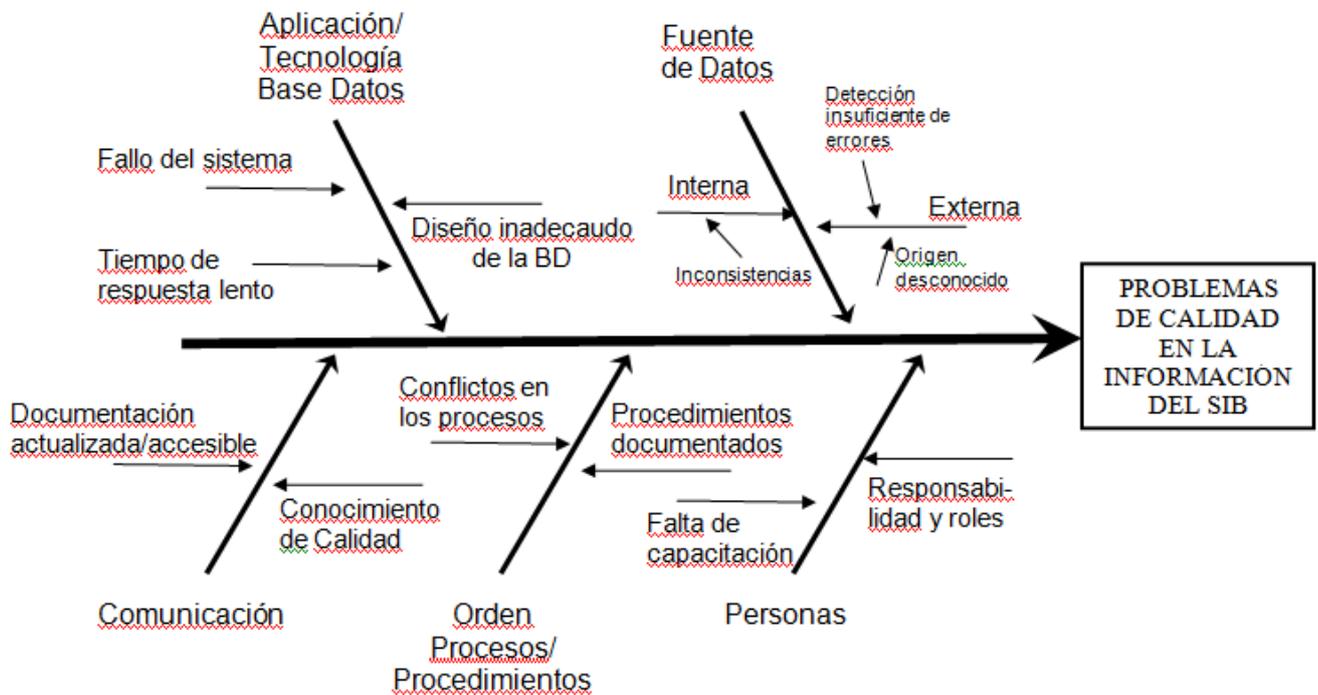
3.1.5 Etapa 5. Identificar causas raíz

Se estableció una línea base para recolectar información mediante la tormenta de ideas, sintetizando un conjunto de datos verbales (ideas, opiniones, temas, expresiones) agrupándolos en función de la relación que tienen entre sí (categorías) y llegar al punto que se describan en términos de qué, cómo, cuándo, dónde, quiénes, etc., para con posterioridad elaborar un cuadro completo de todas las posibles causas y validarlas.

La herramienta utilizada en la determinación de las causas es el diagrama causa-efecto, donde una vez definidas las causas de los problemas desglosados se dejaron en evidencia dónde se encontraban las raíces. La determinación del uso de la herramienta fue la necesidad de estructurar la actividad y los beneficios que aporta: asegurar que las mayores causas no se ignoren permitiendo la completa participación del personal involucrado, que los enfoques en el lado humano y medioambiental así como los técnicos fueran los primeros candidatos para llevar a cabo la prevención de las causas y obtener datos adicionales que rodean al problema para aseverar que no se omitieron otras causas a través de asunciones incorrectas.

Las principales categorías que contribuyeron al efecto o problema fueron:

1. Aplicación/Tecnología Base Datos
2. Fuente de Datos
3. Comunicación
4. Orden Procesos/Procedimientos
5. Personas



Elaborar el diagrama fue una labor educativa en sí misma, que favoreció el intercambio de técnicas y experiencias, ayudando a determinar si los factores seleccionados fueron realmente las causas del problema. El análisis e interpretación del diagrama resultante mostró la habilidad profesional del personal encargado del proceso para eliminar las causas que no aplicaban. Se examinó el “equilibrio” del diagrama comprobando niveles comparables de detalles para todas las categorías. A continuación los factores que se exploraron:

- Un agrupamiento de causas secundarias en una de las categorías (Fuente de Datos) que indicó la necesidad de un estudio más profundo,
- Una categoría principal (Personas) con sólo una causa específica que mostró la necesidad de identificar posibles causas futuras,
- Causas que aparecen repetidamente (documentación, capacitación, roles y habilidades) y la urgencia de cuantificar los efectos de cualquier cambio que se haga.

Más pretenciosamente, se condujo algunas experiencias para verificar, identificar y acorrallar las causas raíces en las que se puede asumir la acción inmediata.

3.1.6 Etapa 6. Plan de mejoras

Para el éxito en la implementación del plan de mejora un elemento importante lo constituye la capacitación de directivos, especialistas y usuarios, acción que se recomienda sea guiada por los miembros del equipo de trabajo. Su objetivo es transmitir los conocimientos sobre la necesidad de la buena calidad de los datos y lo relacionado con la aplicación del procedimiento estableciendo las recomendaciones para la acción.

Los planes de mejoras derivados de la aplicación del procedimiento involucran cambios en el sistema operacional y promueven el correcto diseño de los procesos (validación, actualización) y flujos de datos teniendo en cuenta no sólo el adecuado uso de la tecnología informática, sino también la correcta selección del personal que produzca una o más alternativas para gobernar la reingeniería en la mejora de los flujos y los procesos. Al mismo tiempo de evaluar labores preventivas de futuros errores, una campaña de limpieza y corrección de datos al repositorio fue necesaria y facilitada por la previa determinación de las causas y problemas fundamentales de los datos para establecer los métodos y herramientas a emplear y posteriormente reemplazar los datos “sucios” originales. Se propone trabajar el perfil de datos de forma periódica, así como otros aspectos que minimicen la posibilidad de ocurrencia de errores, como los cambios a la aplicación y modificar la base de datos para dar soporte a la calidad de la información.

3.1.7 Etapa 7. Plan de comunicación

Para todas las etapas del procedimiento se concibe la comunicación como el arranque a las nacientes habilidades que revelan cualquier información esencial. Poner el plan al día a lo largo del proyecto ha sido útil como herramienta de referencia para recordar, notificar y documentar los esfuerzos de comunicación global.

La comunicación ha sido guiada de dos maneras: asegurar incluir las acciones y medios de comunicación para obtener la retroalimentación de la audiencia y crear la oportunidad para el diálogo relacionando las preguntas con las preocupaciones. Los miembros del equipo proponen agregar un campo que la plantilla no contemplaba inicialmente: el estado (con relación a la actividad o etapa del procedimiento) en que se encuentra la acción o mensaje, que permite darle mejor seguimiento.

Audiencia	Mensaje y acción	Medio de Comunicación	Desarrollo y entrega	Preparación de la acción	Fecha planificada	Fecha Completada	Estado

3.2 Aplicabilidad de las “herramientas” y productos propuestos

Durante la aplicación del procedimiento en el actual SIB quedó demostrado que actividades lógicamente planteadas como parte del proceso de validación deben ser cambiadas, siendo necesario verificar primeramente el origen de la fuente de datos para evitar intromisiones, luego comprobar el orden de campos o atributos para garantizar la obediencia de las reglas de datos por la lógica de programación de la aplicación e incorporar reglas del negocio ocultas en los datos que han sido descubiertas aplicando perfiles.

La herramienta propia para la extracción de los datos de la fuente operacional, la transformación y carga en tablas auxiliares que responden a la metodología para cada modelo informativo evitan validar formatos y garantizan el orden de los atributos (consumo de tiempo durante la verificación), el uso de restricciones y disparadores para un mayor control. Se pueden definir además, llaves primarias y foráneas entre tablas relacionadas (modelos con informaciones afines).

Los perfiles de datos que focalizan el análisis de las instancias de atributos individuales se realizan con la herramienta de código abierto *Power*Architect* del grupo *SQL Power*, cuyos resultados permiten ser guardados en los formatos: CSV, PDF o HTML y visualizados en tablas o gráficos. Para el perfil de dependencias y perfil a través de tablas con el propósito de descubrir inconsistencias e incompatibilidades, se utilizó *MS Visual Studio Integration Services*, la tarea genera un archivo XML (*eXensible Markup Language*) de salida y se analizan los resultados con la herramienta *Data Profile Viewer*. Por otra parte las solicitudes de perfiles dependen de la cantidad de datos existentes en la fuente temporal de datos, el número de solicitudes y la configuración seleccionada.

El análisis detallado de los resultados en ocasiones no resultó suficiente. Los archivos generados en formato XML pueden incorporarse a bases de datos simplemente utilizando

SSIS (*SQL Server Integration Services*) como consecuencia puede interesar, por ejemplo, almacenar un histórico de los perfiles para poder evaluar cómo ha ido mejorando la calidad de los datos tratados.

Los gastos en la aplicación del procedimiento están dados por el gasto de entrenamiento para su aplicabilidad, el uso y comprensión de las herramientas propuestas, así como el esfuerzo y tiempo para implementar la captura y evaluación de los datos con la herramienta propia, ya que no se requiere realizar inversiones en otros recursos.

3.3 Conclusiones parciales

La aplicación del procedimiento al repositorio del SIB de muchos años donde la calidad de los datos se deteriora demuestra que:

- Varios procesos afectan la eficiencia de la calidad de los datos. Estos pueden ser agrupados en tres categorías de alto nivel: procesos que producen datos externamente (conversión de datos origen, entrada manual de datos, incorporación de datos por lotes, falta de sincronización durante el intercambio y actualización de datos en tiempo real), los procesos que causan detrimento natural de los datos (reglas no capturadas, nuevos datos, cambios en el sistema operacional, automatización de procedimientos, pérdida del conocimiento) y finalmente los procesos que cambian los datos internamente (procesamiento de transacciones reglamentadas, riesgos y complejidades de limpieza automatizada, depuración).
- Tres componentes proveen la certeza de la calidad de los datos de manera decisiva: las dimensiones de calidad que fueron dirigidas, la metodología para ejecutar las actividades, y por último, la manera en que el grupo de analistas y expertos pueden involucrarse en las actividades.
- Se demostró que la integración de las mejores prácticas y herramientas, asegura las políticas (objetivos, estrategias, metas) que han sido desarrolladas en el procedimiento a disposición de la organización y facilita una adecuada gestión de la calidad de la información en el SIB.

CONCLUSIONES Y RECOMENDACIONES

La calidad de los datos está adquiriendo una progresiva connotación internacional y nacional en aras de disminuir costos, mejorar la imagen de las empresas u organizaciones y la satisfacción de los clientes. El análisis de esta función, desde su perspectiva más contemporánea para el SIB fue centrado en dos ideas fundamentales, que permiten ratificar la hipótesis formulada: el proceso de valoración y análisis de las características de los datos que son el fundamento para elevar el conocimiento de la calidad de la información en el SIB y el proceso de mejora con un enfoque proactivo que fomente la eliminación de datos defectuosos.

Con la investigación realizada se logra concluir que resultó necesario elaborar un procedimiento de gestión de la calidad de datos financieros en el SIB con un enfoque sistemático, formal y preciso a través de los conocimientos adquiridos y sobre todo efectivo mostrando la validez de la propuesta y justificación de su uso, con el objetivo de garantizar la integridad de los datos del SIB, permitiendo adicionar y mantener en el repositorio solamente información confiable y de alta calidad.

Durante el desarrollo del trabajo, estas conclusiones ponen de relevancia aspectos claves que se resumen en las siguientes afirmaciones:

- Existe una creciente base teórica conceptual sobre el tema de la calidad de datos, que se ha ido enriqueciendo paulatinamente a partir de contribuciones teóricas. Sin embargo, en contraste con el desarrollo empírico y teórico conceptual, en la componente metodológica de la forma de entender, orientar y ejecutar sus actividades, se han apreciado escasos procedimientos que permitan dotar al BCC de una guía para la definición y gestión de la calidad de los datos financieros, desde un enfoque proactivo y estratégico. Por lo cual el problema científico formulado para la presente investigación se considera de gran actualidad y pertinencia, tanto en el plano conceptual – metodológico como práctico.
- El procedimiento general propuesto y las etapas asociadas conforman un cuerpo de elementos coherentes desde la perspectiva teórico-metodológica desarrollada para

dar solución al problema planteado, constituye también un instrumento apreciable con las pautas y “buenas prácticas” que potencia de manera efectiva la eliminación de las causas de información defectuosa, permite desarrollar e implementar adecuadamente las actividades de gestión de la calidad de una forma disciplinada, coherente y repetitiva en virtud de lograr y mantener la calidad de los datos en el SIB. Se demostró que la integración de técnicas de perfil de datos y la herramienta propietaria implementada facilitan la realización de las tareas y actividades en la gestión de la calidad con un mejor desempeño.

- La aplicación del procedimiento desarrollado puso en evidencia las deficiencias en la definición del esquema de la base de datos, la necesidad de mejorar los procesos actuales, las causas de los problemas de calidad de los datos en el SIB por la falta de conciencia sobre la importancia y conocimiento del tema. Además, permitió obtener la especificación de calidad de datos formal y estructurada con esfuerzo planeado y permanente de todos los participantes para la satisfacción de los diversos departamentos del BCC.

El trabajo deja abiertos algunos aspectos importantes como la extensión del procedimiento al uso de otras técnicas y herramientas para la mejora evolutiva de la calidad de datos en el Sistema Bancario y la incorporación de una nueva etapa para la implementación de controles que garantice la retroalimentación, a través de la cual, se mida el actual desempeño real comparando con lo estandarizado y actuando en las diferencias para que la calidad no sea un proceso de inspección y monitoreo, sino más bien de prevención de errores.

Como resultado del estudio realizado, así como de las conclusiones generales emanadas del mismo, se recomienda:

- Continuar la divulgación de las experiencias y el resultado obtenido en este trabajo de investigación, a través de publicaciones en la revista del BCC y en el próximo evento científico de Eficiencia Bancaria Nacional, todo lo cual contribuirá a la difusión y conclusión de dicho resultado.
- Establecer una valoración periódica de la calidad de los datos mediante la incorporación de los resultados de los perfiles de datos en un proyecto *Business*

Intelligence que permita elaborar informes e indicadores sobre la variación de la calidad de los datos.

- Instituir la entrada de datos de fuentes externas en formato XML, para determinar si el modelo informativo se ciñe a las restricciones descritas en el esquema utilizado para su construcción y reducir el costo computacional en el actual proceso de validación del SIB. Controlar el diseño de documentos a través de esquemas aumenta el grado de fiabilidad, consistencia y precisión.

BIBLIOGRAFIA

- Abate, M. & Diegert, K. (1998). A Hierarchical Approach to Improving Data Quality. Recuperado el 2011, de [http:// www.dataquality.com](http://www.dataquality.com)
- Acuña, E. & Rodríguez, C. (2009). The treatment of missing values and its effect in the classifier accuracy: Four different methods to deal with missing values.
- Akoka, L., Berti-Équille, O., Boucelma, M. & Buzeghoub, M. (2007). A Framework for Quality Evaluation in Data Integration Systems. 9th International Conference on Enterprise Information Systems (ICEIS'2007), Funchal, Portugal.
- Ballou D. & Pazer H. (1995). Modeling Data and Process Quality in Multi-Input, Multi-Output Information Systems. *Management Science* , 31.
- Barnett, V. & Lewis, T. (1994). *Outliers in statistical data* (3rd edition ed.). John Wiley & Sons.
- Batini, Carlo & Scannapieca, Monica. (2006). *Data Quality-Concepts, Methodologies and Techniques*. Springer. ISBN/13978-3-540-33172-8.
- Benítez, G. (2005). *Calidad de Datos, factor crítico para la administración de los recursos*. Recuperado el 3 de Febrero de 2011, de http://www.pyme.com.mx/articulos_pyme/todoslosarticulos/calidad_de_datos_factor_critico.
- Bright, L. & Raschid, L. (2002). Using Latency-Recency Profiles for Data Delivery on the Web. Of the 28th International Conference on Very Large Databases.
- Brodie.M., Mylopoulos, J. & Schmidt,J. (1984). *On the Development of Data Models. On Conceptual Modelling*. Springer Verlag.
- Calero, C., Caballero, I., Serrano, M. y Piattini, M. (2003). *Calidad de los Almacenes de Datos*. España.
- Chisholm, M. (2001). *Managing Reference Data in Enterprise Databases: Binding Corporate Data to the Wider World*. Morgan Kaufmann.

- Chisholm, M. (2004). *How to Build a Business Rules Engine: Extending Application Functionality Through Metadata Engineering*. Morgan Kaufmann.
- Dataqualitypro. (2010). The Essential Online Resource for data quality professionals. *Data Quality Articles Journal* .
- De Amicis, Fabricio & Batini, Carlo. (2004). A Methodology for Data Quality Assessment on Financial Data. *Studies in Communication Sciences* 4/2 , 115-136.
- Deming, W. Edward. (2000). *Out of the Crisis*. MIT Press.
- Dylan, Jones. (January 2009). Resources for Creating a Data Quality Methodology. Recuperado el 12 de Enero de 2010, de http://www.dataqualitypro.com/?page=DQ_method_resources
- Dylan, Jones. (August 2009). Interview with Larry English, Creator of TIQM (Total Information Quality Management). Recuperado el 3 de Marzo de 2012, de <http://www.dataqualitypro.com>
- Dylan, Jones. (2009). *Data Quality Book Review: Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information*. Recuperado el 11 de Octubre de 2010, de http://www.dataqualitypro.com/?page=ten_steps_review
- Dylan, Jones. (August 2009). Interview with Danette author of *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information*. Recuperado el 12 de Marzo de 2012, de http://www.dataqualitypro.com/?page=danette_mcgilvray_i1
- Dylan, Jones. (November 2009). *Productivity Tools for the Innovative Data Quality Leader*. Recuperado el 14 de Noviembre de 2011, de http://www.dataqualitypro.com/?page=prod_tools_leaders
- Dylan, Jones. (August 2010). *Ten Techniques for Data Quality Root-Cause Analysis*. Retrieved Febrero 22, 2012, from <http://www.dataqualitypro.com/>
- Dylan, Jones. (March 2012). *How to Transition from Quality Management to Data Quality Management: Interview with Gabriel Gómez Mantilla, General Manager of Wolman Group*. Recuperado el 12 de Marzo de 2012, de www.dataqualitypro.com.
- English, Larry. (October 1996). Help for data quality problems - A number of automated tools can ease data cleansing and help improve data quality. *Information Week* , 53.

- English, Larry. (1999). *Improving data Warehouse and Business information Quality*. John Wiley & Sons.
- English, Larry. (September 2003). Total information quality management – A complete methodology for IQ management. (D. R. Magazine, Ed.) Retrieved from <http://www.dmreview.com>
- English Larry. (2007). IQ/DQ Glossary (IAIDQ). Recuperado el 5 de Enero de 2010, de <http://www.iaidq.org/main/glossary.shtml>
- Eppler, M. J. (2003). *Managing Information Quality: Increasing the Value of Information in Knowledge-intensive Products and Processes*. Springer.
- Escalona, M.J., Torres, J., & Mejías, M. (2002). *Requirements capture workflow in Global Information Systems*. Montpellier, France: Springer-Verlag.
- Etcheverry, L., Gatto, P.y Tercia, S. (2005). *Análisis del proceso de carga de Data Warehousing de Enseñanza de la Facultad de Ingeniería*. Montevideo, Uruguay.
- Etcheverry, L., Tercia, S., Marotta, A. y Peralta, V. (2007). *Medición de la Exactitud de Datos en Sistemas Fuentes: Un Caso de Estudio*. Facultad de Ingeniería, Instituto de Computación. Universidad de la República.Uruguay.
- Fayyad U., Piatetsky-Shapiro & Smyth P. (1996). *From Data Mining to Knowledge Discovery in Databases*. Recuperado el 13 de Mayo de 2008, de AI Magazine, AAAI: <http://www.aaai.org/AITopics/assets/PDF/AIMag17-03-2-article.pdf>
- Fernández, V. J. (1999). *Las Siete Herramientas Básicas de la Calidad*. Recuperado el 23 de Enero de 2012, de <http://asesoriajuridicagratis.blogspot.com/>
- Gertz, M., Ozzu, M., Sattler, K. & Saake G. (2004). Report on the Dagstuhl Seminar: Data Quality on the Web. SIGMOD Record , 33.
- Guerrero, Julio. (2006). La Eficiencia en las estadísticas y el Sistema Informativo Bancario. Revista del Banco Central de Cuba. ISSN/1560-795X , 9(3), 22-24.
- Gutiérrez, H. y De la Vara, R. (2004). *Control Estadístico de la Calidad y Seis Sigma*. México: McGraw-Hill.

- Hay, David C. (2003). What Exactly is a Data Model. Retrieved December 26, 2009, from www.dmreview.com/issues/20030201/6281-1.html
- Hay, David C. (2006). *Data Model Patterns: A Metadata Map*. Morgan Kaufmann.
- Heredia, J. y Vilalta J. (2009). La calidad de los datos: Su importancia para la gestión empresarial.
- Jarke, M., Jeusfeld, M.A., Quix, C. & Vassiliadis, P. (1999). Architecture and Quality in Data Warehouses: An Extended Repository Approach. *Information Systems* , 24(3).
- Juran, J. M. (1995). *Managerial Breakthrough: The Classic Book on Improving Management Performance*. McGraw Hill.
- Levy, S. (2004). Model Documents and forms for Organizing and Maintaining a Data Quality Program. Retrieved November 10, 2009, from [www .dataqualitymodeldocument.com](http://www.dataqualitymodeldocument.com)
- López, B. y Pérez, R. (2002). ¿Tiene usted datos sucios?. GIGA (La Habana, Cuba).
- López, B. y Pérez, R. (2007). Identificación automatizada de errores en las bases de datos, del DBAnalyzer 1.0 al DBAnalyzer 2.0. Memorias del III Congreso Internacional de tecnologías, contenidos multimedia y realidad virtual "Informática 2007" , ISBN/978-959-286-002-5 .
- Loshin, David. (2001). *Enterprise Knowledge. Management: The Data Quality Approach by Chapter 5: Dimensions of Data*. Recuperado el Marzo de 2011, de unjobs.org/authors/david-loshin
- Loshin, David. (2001). *Enterprise Knowledge Management: The Data Quality Approach*. Morgan Kaufmann. ISBN/0124558402, 9780124558403.
- Loshin, D. (2001). *Integration and the Data Quality Imperative: The Data Quality Monitor*. Retrieved January 12, 2012, from [http:// www.datajunction.com](http://www.datajunction.com).
- Loshin, David. (2009). *Data Quality Management Software Product Directory 2009 EDITION*. Recuperado el 9 de Marzo de 2011, de SearchDataManagement: www.mdmbook.com
- Loshin, David. (2009). *Master Data Management Platforms Directory 2009 EDITION*. Recuperado el 9 de Marzo de 2011, de SearchDataManagement.: <http://cdn.ttgtmedia.com/searchDataManagement/downloads/SDMMasterDataDirectory2009.pdf>

- Loshin, David. (2006). Monitoring Data Quality Performance Using Data Quality Metrics.
- Mark, Mosley. (2008). Guide DAMA-DMBOK Functional Framework. Recuperado el 1 de Marzo de 2012, de http://www.dama.org/files/public/DAMADMBOK_Functional_Framework_v3_02_20080910.pdf
- Marotta, A. & Ruggia, R. (2005). Managing Source Quality Changes in Data Integration Systems. Uruguay.
- Martirena, E. (2008). Medición de la calidad de datos: Un enfoque parametrizable. Recuperado el 28 de Febrero de 2012, de http://www.fing.edu.uy/inco/pedeciba/bibliote/cpap/tesis_martirena.pdf
- Matus, C. (2007). Dimensiones de la Calidad según OECD y EUROSTAT. Departamento Investigación y Desarrollo Instituto Nacional de Estadísticas de Chile.
- Maydanchik, A. (2007). Data Quality Assessment by chapter Thirteen causes of enterprise data quality problems. (T. Publications, Ed.) Retrieved January 23, 2012, from <http://searchdatamanagement.techtarget.com/feature/Thirteen-causes-of-enterprise-data-quality-problems>.
- McGilvray, Danette. (2008). Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information. ScienceDirect. ISBN/978-0-12-374369-5.
- Montero, L. (2004). Visión de la Calidad de Datos. IBM- Business Consulting Services.
- Motro, A. & Rakov, I. (1998). Estimating the quality of databases. Of the 3rd International Conference on Flexible Query Answering Systems. Roskilde, Denmark.
- Naumann, F., Freytag, J. & Leser, U. (2003). Completeness of Information Source.
- Olson, Jack E. (2003). Data Quality: The Accuracy Dimension. Morgan Kaufmann. ISBN/1558608915.
- Pande, Peter S., & Holpp Larry. (2002). What Is Six Sigma?. McGraw Hill.
- Peralta, V. (2006). Data Quality Evaluation in Data Integration Systems. PhD Thesis, Université de Versailles – France and Universidad de la República – Uruguay.

- Peralta, V., Ruggia R. & Bouzegoub, M. (2004). Analyzing and Evaluating Data Freshness in Data Integration Systems. *Ingénierie de Systèmes d'Information (ISI)* , 9 (5/6), 145-162.
- Pérez, D. (2008). Análisis, Rediseño y Aplicación del Procedimiento de Diagnóstico de la Calidad de los Datos. Trabajo de Diploma, Cujae, La Habana.
- Pérez, D. (2010). Rediseño e implementación del Procedimiento de Diagnóstico de la Calidad de los Datos. Cujae, La Habana.: Tesis presentada en opción al título de Máster en Calidad Total.
- Pérez, D. y Vilalta J. (2009). El Procedimiento de Diagnóstico de la Calidad de los Datos y el uso de las tecnologías de la información. Evento Semana Tecnológica. ISSN/2076-9792.
- Pérez, D. y Vilalta, J. (2010). Aplicación del Procedimiento de diagnóstico de la Calidad de los Datos en empresa productora de envases de madera. Recuperado el 21 de Marzo de 2012, de http://dialnet.unirioja.es/servlet/fichero_articulo?codigo=3656603
- Pérez, D y Vilalta, J. (2010). El procedimiento de diagnóstico de la calidad de los datos con enfoque de riesgos. Recuperado el 28 de Marzo de 2012, de <http://semanatecnologica.fordes.co.cu/ocs-2.3.2/public/site/414.pdf>
- Pipino, L., Yang W. Lee, and Wang Richard. (2002). Data Quality Assessment at Communications of the ACM.
- Porras, Mínguez A. (Septiembre 2008). Fundamentos de la Calidad de Datos Aules d'Empresa. Recuperado el 27 de Octubre de 2011, de http://seminarisempresa.fib.upc.edu/anteriors/2008/programes/PWD_DWH_S2_FundamentosCalidadDatos.pdf
- Rahm, & H, D. (2000). Data Cleaning: Problems and Current Approaches. *Data Engineering Bulletin* , 23(4), 3-13.
- Redman, Thomas C. (2001). *Data Quality: The Field Guide*. Digital Press.
- Redman, Thomas C. (2001). *Sistemas de calidad de datos de segunda generación. Manual de Calidad de Juran*. McGraw Hill.
- Redman, Thomas C. (2004). Data an unfolding quality disaster. Recuperado el 10 de Noviembre de 2011, de <http://www.dmreview.com/portals/dataquality>

- Ross, R.G. (2005). Business Rule Concepts: Getting to the Point of Knowledge. Business Rule Solutions.
- Ross, R.G. (2006). Los Principios de la Independencia de las Reglas. Recuperado el 8 de Octubre de 2009, de <http://www.businessrulesgroup.org/brmanifesto/BRManifiesto%5Bv1%5B1%5D.0%5D.pdf>
- Schober. (2009). Servicios y Soluciones de Schober para la calidad de Bases de Datos. Recuperado el 4 de April de 2009, de <http://www.Schober.es/bases-datos.schober.es/calidad/SchoeberDataQuality.pdf>
- Scofield, M. (February 2008). Fundamentals of Data Quality. NoCOUG Journal .
- Silverman, L. (2006). Wake Me Up When the Data Is Over: How Organizations Use Stories to Drive Results. Jossey-Bass.
- Tan, P. S. (2006). Introduction to datamining: Missing values. 40-41.
- Taylor, J. & Cihon, C. (2004). Statistical techniques for data analysis (2 ed.). Chapman & Hall/CRC.
- Tierstein, L. (2005). A Methodology for Data Cleansing and Conversion. (L. White paper W R Systems, Ed.)
- Trillium, Software. (2004). Ten Critical Factors for Successful Enterprise-Wide Data Quality. (Trillium Software™) Recuperado el 8 de Octubre de 2009, de www.trilliumsoftware.com
- Trillium, Software. (2004). How Data Profiling & Analysis saves Companies \$millions. (Trillium Software™) Recuperado el 22 de Octubre de 2009, de www.trilliumsoftware.com
- Uniserv. (2005). Soluciones de Software al servicio de la calidad de datos. Retrieved October 12, 2009, from <http://www.uniserv.com>.
- Uribe, I. (2010). Guía metodológica para la selección de técnicas de depuración de datos. Recuperado el 12 de Enero de 2012, de Universidad Nacional de Colombia: <http://www.bdigital.unal.edu.co/2033/1/71644758.20101.pdf>
- Vidaña, R. R. (2010). Siete Herramientas Informáticas Básicas para la Calidad. Recuperado el 26 de Abril de 2012, de www.capasitio.com

- Vilalta, J. (2008). Procedimiento para el Diagnóstico de la Calidad de los Datos en organizaciones cubanas. Cujae, La Habana: Tesis presentada en opción al grado científico de Doctor en Ciencias Técnicas.
- Wang, Hai & Wang, Shouhong. (2009). Rough set analysis for total Information Quality Management. Recuperado el 9 de Marzo de 2012, de <http://ojs.acadiau.ca/index.php/ASAC/article/viewFile/567/476>
- Wang, R.& Strong, D. (1996). Beyond accuracy: What data quality means to data consumers. 12 (4), 5-34.
- Wikipedia. (2009). Data Integration, Data Mapping, Data Model, DataStore, Data Transformation. Retrieved Diciembre 18, 2009, from http://en.wikipedia.org/wiki/Main_Page.
- WILSHIRE & DAMA. (2007). Learn about Introduction to Data Quality Tools and Technologies. WILSHIRE Meta-Data Conference & DAMA International Symposium– March 4-8.
- Zelaya Fajardo, J. (Enero 2010). Control de la Calidad Total: Herramientas Administrativas de la Calidad Total. Recuperado el 10 de Abril de 2012, de <http://aplicaciongr.blogspot.com/2007/02/las-siete-nuevas-herramientas-de-la.html>

ANEXOS

Anexo 1. Modelos informativos del SIB.

Código	Grupo	Nombre del Modelo	Entrega M- mensual S- semanal T- trimestral E- eventual DD –días
0101	01	Cuentas por cobrar en MLC (en Cuba y/Exterior)	M20
0102	01	Cuentas por cobrar en MN	M20
0103	01	Desglose Cuentas x Cobrar por Provincias	M30
0199	01	Cuentas por Cobrar en MLC y MN (1999 y 2000)	
0201	02	Ctas y Efectos x Pagar a Exterior-Proveedores - Corto Plazo	M15
0202	02	Pagos al Sector Privado y Cooperativo - Cheques Voucher	M15
0203	02	Cuentas Por Pagar en MN (Global)	M20
0204	02	Cuentas Por Pagar en CUP - Sector Agropecuario NO Estatal	M20
0300	03	Financiamiento Bruto Concedido en CUP, CUC y Otras divisas	M07
0301	03	SalDOS Pendientes de amortizar en CUP, CUC y Otras divisas	M07
0400	04	Cuentas en CUC y otras divisas en Bancos cubanos	M07
0401	04	Ctas de Personas Juridicas en CUC y Otras divisas	M10
0402	04	Flujo de Entradas y Salidas divisas con el exterior	E05
0403	04	Cuentas de Personas Juridicas en CUP	M12
0501	05	Medios de Pago del Sistema Bancario Nacional	M10
0517	05	Circulacion Monetaria x OACE	M14
0518	05	Circulacion Monetaria x OACE. AHORRO	M15
0600	06	Encaje Legal	
0701	07	Cuentas de Clientes en MN, CUC y Otras divisas (Pers Naturales y Juridicas)	M12
0702	07	Depósito a Plazo Fijo en MN de Entidades en Perfeccionamiento	E20
0703	07	Cuenta de Particulares en CUC y otrs Divisas - Mensual	M07
070A	07	Ctas. de Particulares en CUC y Otras Divisas Semana 1	S01
070B	07	Ctas. de Particulares en CUC y Otras Divisas Semana 2	S02
070C	07	Ctas. de Particulares en CUC y Otras Divisas Semana 3	S03

070D	07	Ctas. de Particulares en CUC y Otras Divisas Semana 4	S04
070E	07	Ctas. de Particulares en CUC y Otras Divisas Semana 5	S05
0910	09	Financiamientos vencidos, renegociados o reestructurados	M07
0911	09	Cheques Deficientes - Personas Juridicas	M10
0914	09	Cheques Deficientes - Personas naturales y otros	M10
0915	09	Letras de cambio no pagadas o pagadas parcialmente	M10
0916	09	Transacciones rechazadas y Medidas aplicadas a infractores	M10
0917	09	Sentencias Pagadas y Pendientes de Pago	M12
0920	09	Balance General. Supervisión a Distancia	M05
0921	09	Estado de Resultado. Supervisión a Distancia.	M05
0922	09	Cuentas por Fideicomiso. Balance General.	M10
0923	09	Cuentas por Fideicomiso. Estado de Resultado.	M10
1001	10	Pasivos Externos - Deudas al Cierre del mes	M15
1002	10	Pasivos Externos - Flujos (movimientos) del mes	M15
1003	10	Pasivos Externos-Sector Financiero-Proyección de Pagos	M05
2001	20	Activos Externos - Saldos al Cierre del mes	M15
2002	20	Activos Externos - Flujos (movimientos) del mes	M15
2400	24	Cuentas corrientes de los OACE en CUP, CUC y Otras divisas	M12
2401	24	Financiamiento y Cartera de Prestamos MN	M15
2500	25	Servicios y Transferencias con el exterior - Balanza de Pagos	T20
2600	26	Tasas Activas y Pasivas en CUP, CUC y Otras divisas	E15
2701	27	Tipo de Cambio	M05

Anexo 2. Productos líderes para la calidad de datos.

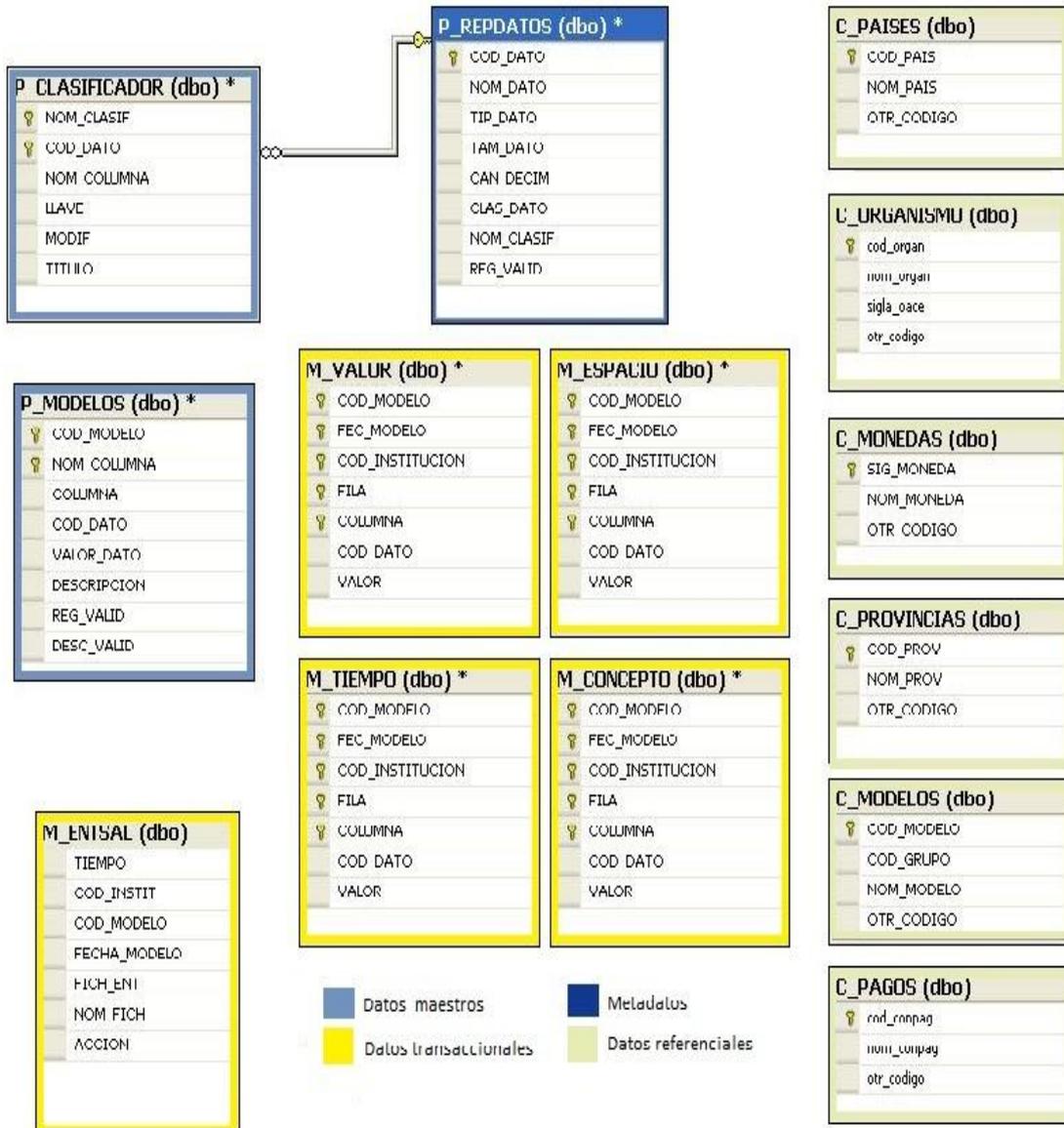
Vendedor	Producto	Alcance
Business Objects (now SAP)	Business Objects Data Quality, Universal Data Cleanse, Business ObjectsData Quality for SAP, Business Objects Data Quality for Oracle's Siebel CRM, Business Objects Data Insight	Productos múltiples para la calidad de datos, monitoreo, análisis, estandarización y reportes.
DataFlux	DataFlux Data Quality Integration Platform	Productos múltiples soportando diferentes procesos de calidad de datos en una única plataforma.
DataLever	DataLever Enterprise Suite	Es un proceso visual diseñado para crear, ejecutar transformaciones de alto rendimiento y procesos de calidad de datos. Incluye reingeniería y perfil de datos.
DataMentors Inc.	DataFuse	Es un sistema que prepara e integra una base de datos modular que limpia, organiza y prepara datos.
Eprentise	Eprentise Data Quality	Es un software que estandariza, localiza y resuelve datos duplicados.
Harte-Hanks Trillium Software	TS Insight, TS Discovery, TS Quality, TS Enrichment	Es una colección de softwares de gestión integrada de calidad. global
IBM	IBM Information Server WebSphere Quality Stage	Proporciona la calidad de los datos para el manejo de direcciones de clientes (estandarización e identificación de duplicados).
Informatica	Informatica Data Quality	El paquete incluye posibilidades de identificación de duplicados, verificación de emails y validación en base a diccionarios y reglas del negocio. Permite identificar problemas de completitud, conformidad, consistencia, integridad y exactitud.
Innovative Systems Inc.	i/Lytics Enterprise Data Quality Suite: i/Lytics Data Profiler, i/Lytics Data Quality, i/Lytics GLOBAL	Realiza limpieza y reconciliación de datos, detección de duplicados y una gestión de cambios que asegure la sincronización de los datos. Licencia una edición real-time y otra batch.
Netrics	Netrics Matching Platform	Incluye cuatro componentes: permite encontrar correspondencias difusas, crear modelos de decisión utilizando técnicas de aprendizaje automático para encontrar duplicados e identificar relaciones, reportes y reconciliación de datos.
Oracle Corp.	Data Quality for Oracle Data Integrator	Incluye múltiples herramientas para la calidad y gobierno de datos. Ofrece solución de MDM

		especializada en datos financieros y para la toma de decisiones. Incluye interfaces gráficas, e integración con herramientas de workflow mediante Web Services.
Talend	Talend Open Profiler	Abierta a fuentes de perfil de datos. Define estadísticas y métricas.
Uniserv	Data Quality Batch Suite	Software de calidad de datos y servicios particularmente para inteligencia de negocio.
Zoomix (now Microsoft) MICROSOFT SUBSIDIARY)	Zoomix Accelerator	Resuelve dinámicamente y en tiempo real cualquier inconsistencia, valiéndose de tecnologías de análisis semántico y lingüístico, así como aprendizaje automático.
Microsoft	SQL Server Integration Services	Plataforma de integración de datos con un motor de procesamiento paralelo que coordina el flujo de control entre las tareas dentro de SSIS y un motor de flujo de datos para las tareas de transformación y procesamiento.
SQL Power	Power MatchMaker, Power Architect	Herramientas desarrolladas en Java (multiplataforma). Ataca varios problemas de limpieza y perfil de dato

Anexo 3. Productos líderes en MDM.

Vendedor	Producto	Alcance
DataFlux	DataFlux qMDM	Diseñada para incorporar tecnología demostrada que consolide y maneje inteligentemente los datos. Entrega una vista más unificada de la empresa a través de las aplicaciones y plataformas.
Data Foundations ATA FOUNDATIONS	OneData	Ofrece un uso comprensivo, de la funcionalidad del ciclo vida para los datos maestros junto con el apoyo para múltiples estilos MDM y acercamientos arquitectónicos.
Enterworks	Enterworks Enable MDM	Maneja la dirección entre las fuentes de los datos para descubrir valiosas relaciones entre los juegos de datos y hace el trabajo más eficazmente para alcanzar las metas de la organización.
FullTilt Solutions	Perfect Product Suite	Automatiza y maneja el completo y exacto proceso de sincronización de datos internos para habilitar el negocio-a-negocio perfeccionado con las capacidades del workflow automatizadas.
IBM	IBM InfoSphere Master Data Management Server	Diseñado para el manejo crítico de los datos maestros a través de clientes, productos y cuentas de dominios provistas desde aplicaciones Back-Office a través de múltiples interfaces.
Kalido	Kalido Master DataManagement	Proporciona una sola plataforma para describir, armonizar y gobernar los datos del multi-dominio, para soportar inteligencia de negocio exacta y la gobernación de datos.
Microsoft	SQL Server 2008 R2 Master Data Services	Habilita a las organizaciones comenzar con herramientas simples para requisitos operacionales y analíticos, adapta el software a los requisitos adicionales incrementalmente.
Oracle Corp.	Oracle Customer Hub	Software para centralizar la información de sistemas heterogéneos, creando una sola vista de información que pueda apalancar los departamentos funcionales y los sistemas analíticos.
SAP	SAP NetWeaverMDM	Integra y soporta los procesos de negocio a través de la cadena de valores; consolida, sincroniza, distribuye, centraliza, administra datos maestros y habilita catálogos de búsquedas.

Anexo 4. Esquema del modelo lógico del SIB.



Anexo 5 .Reporte de Errores en el SIB. Modelo a validar: 0201 del MIC (Ministerio de Informática y las Comunicaciones)

Fichero o Modelo	Columna	Fila	Error
V201MIC.DBF	IMP_NFREC	0	Error en la definición del nombre, el tipo ó el tamaño de este campo.
V201MIC.DBF	EXT_EFECTO	0	Error en la definición del nombre, el tipo ó el tamaño de este campo.
V201MIC.DBF	IMP_INTMOR	0	Error en la definición del nombre, el tipo ó el tamaño de este campo.
V201MIC.DBF	EXT_TOT_VE	0	Error en la definición del nombre, el tipo ó el tamaño de este campo.
V201MIC.DBF		0	Error en la posición de los campos o columnas del modelo. Deben aparecer en el siguiente orden: 'COD_ONE','NOM_ENTIDA','PRESTAMIST','SIG_MONEDA','EXT_NO_VEN','EXT_EFECTO','IMP_PRINCI','IMP_INTCOV','IMP_INTMOR','EXT_TOT_VE','IMP_NF REC','PAGO_REAL','IMP_INTPV'