

Gobierno de Datos sobre el Padrón de Contribuyentes de la Provincia de Buenos Aires (Modalidad PROYECTO)

Lic. Mauricio Décima¹ - Lic. Eliseo Palacios²

SIE 2018 - 12º Simposio de Informática en el Estado

Abstract. El presente artículo describe las etapas definidas en la Agencia de Recaudación de la Provincia de Buenos Aires para alcanzar un Gobierno de Datos sobre la información de los contribuyentes de la Provincia de Buenos Aires. El proyecto hace foco principalmente en aplicar un proceso de mejora continua sobre la calidad de los datos del padrón de contribuyentes y todos los impuestos relacionados con le objetivos de establecer un proceso que permita definir políticas de gobierno, controles de calidad, procesos de corrección que permitan mejorar la operación, transferencia, calidad y precisión de los servicios al contribuyente en pos de poner a disposición información con un alto nivel de exactitud, completitud, integridad, coherencia y confiabilidad.

Keywords: Calidad de Datos, Gobierno de datos, SAS Data Connector, Changed Data Capture, Data Cleansing, Quality Data, Data Matching, Objetos Imponibles.

1 Introducción

Para la Administración Tributaria, la disponibilidad de información oportuna y confiable es indispensable para lograr sus objetivos. Los sistemas de información tributaria implementan estrategias de recolección, tratamiento y distribución de la información requerida, y son uno de los factores condicionantes para el éxito de una administración.³

¹ Lic. Mauricio Décima – Gerente de Gestión de Proyectos y Servicios – ARBA

² Lic Eliseo Palacios – Jefe de Departamento de Administración de Proyectos - ARBA

³ <http://www.clei2017-46jaiio.sadio.org.ar/sites/default/files/Mem/SIE/SIE-21.pdf> - Mg Sandra D' Agostino

La información respecto de los objetos imposables de la Provincia de Buenos Aires, relativa a Inmobiliario, Automotor, Ingresos Brutos, Agentes de Recaudación y Embarcaciones, se encuentra almacenada en forma desagregada por impuesto. Cada impuesto conserva datos de titularidad y de domicilios del objeto imponible. El esquema de almacenamiento independiente por impuesto complica conocer la situación fiscal de un sujeto en particular.

Actualmente esta información se encuentra en bases de datos de distintos propietarios sin normalización alguna, donde el espectro de la fuente de los datos es amplio en tecnología y en cuanto a la diversidad de aplicaciones que consolidan la misma.

Hoy en día es indispensable proveer al ciudadano contribuyente acceso a la información exacta, completa, íntegra, coherente y confiable.

El activo más importante de cualquier organización es la información. Según lo que plantea la norma ISO 9000: 2000, la calidad se podría definir como "el grado en el que un conjunto de características inherentes cumple con los requisitos, esto es, con la necesidad o expectativa establecida, generalmente implícita u obligatoria" En este sentido, es menester de la agencia de recaudación de la provincia de buenos aires contar con un sistema de gestión de la información que soporte desde la integración del dato hasta la calidad del mismo, aplicando la capa de gobierno necesaria para la seguridad de la información.

El volumen masivo de información generada durante el curso de un número cada vez mayor de complejas transacciones, combinado con los datos que fluyen de incontables fuentes de terceros, suponen desafíos tremendos para la organización que consume innumerables recursos humanos y tecnológicos por mantener altos niveles de calidad de los datos. Como resultado, la información pobre o "sucia" va impregnando los sistemas a lo largo de toda la Agencia, impactando de forma no positiva a lo largo de todos los procesos productivos de la agencia, desde el desarrollo de estrategias y gestión del rendimiento, a la recaudación y servicio al cliente.

Resolver el problema no es tan fácil como pueda parecer. Los esfuerzos por conseguir un nivel adecuado en la calidad de datos se han visto obstaculizados por reglas de negocio incompletas o conflictivas, diferentes técnicas de calificación y métodos de validación, y requisitos para los datos poco claros que no pueden adaptarse a los retos específicos de la agencia. Para poder mejorar los procesos clave del negocio e impulsar el crecimiento con datos que sean correctos, completos y oportunos, las organizaciones necesitan una

forma flexible, fiable y proactiva de preservar la integridad constante de sus activos de información más críticos.⁴

En función de esto que ARBA, más precisamente la Gerencia General de Tecnología a Innovación, con el apoyo de la dirección ejecutiva, impulsa un proyecto de calidad de datos sobre el padrón de contribuyentes de toda la provincia de Buenos Aires sobre todos los impuestos para mejorar la calidad y precisión de sus servicios al contribuyente brindando información exacta, completa, íntegra, coherente y confiable.

Contar con un modelo conceptual en 3ra forma normal, transaccional y con un proceso de calidad que permita alimentar la nueva generación de aplicaciones contribuirá a mejorar la calidad de servicio que se provee al contribuyente.

Las empresas que le dan importancia a la calidad de sus datos, les permiten obtener beneficios claves para agregar valor al negocio otorgando⁵:

- Minimizar los riesgos en sus proyectos, especialmente en los relacionados con Tecnologías de la Información.
- Ahorro de tiempo y recursos, haciendo un mejor uso de la infraestructura tecnológica y sistemas para explotar su información.
- Toma de decisiones de negocio oportunas, en base a información confiable, validada y limpia.
- Adaptación a estándares o regulaciones internacionales sobre el manejo de información, permitiendo facilidad al momento de ejecutarlas.
- Mejorar la confianza, buenas relaciones e imagen de la institución ante sus contribuyentes.

2 Situación-Problema u Oportunidad

Mediante la Ley N° 13.766 se creó la Agencia de Recaudación de la Provincia de Buenos Aires (Arba), un organismo técnico especializado, dotado de las herramientas necesarias para ejecutar la política tributaria a través de la aplicación, recaudación y fiscalización de los tributos y accesorios dispuestos por las normas legales y la administración del catastro territorial de la provincia de Buenos Aires, entre otras acciones. Los pilares sobre los cuales se crea la Agencia son los siguientes:

- **Misión:** Ejecutar eficientemente la política tributaria mediante la determinación, fiscalización, y percepción de los tributos y accesorios dispuestos por

⁴ <http://www.informationbuilders.es/data-quality> - Information Builders

⁵ <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/368784/introducci-n-a-la-calidad-de-datos-definici-n-control-y-beneficios> - powerdata

las normas legales, administrar el catastro territorial y brindar servicios a otros organismos públicos, promoviendo el cumplimiento voluntario y contribuyendo al bienestar económico y social de la ciudadanía.

- **Visión:** Ser un organismo impecable en la calidad de los servicios y en la administración de las relaciones con el ciudadano necesarias para cumplir con las metas de recaudación, conformado por un equipo profesional y comprometido que actúa en un marco de integridad y de permanente búsqueda de la innovación mediante el uso de tecnología de avanzada.

Estos pilares se sustentan en los siguientes valores:

- Servicio al ciudadano
- Transparencia y no discrecionalidad
- Eficacia y eficiencia en la operación diaria
- Participación y compromiso del personal y de la gestión
- Ética

Se encuentran alineados absolutamente con estos y con sus lineamientos estratégicos:

- Optimizar la recaudación
- Optimizar el sistema de servicios
- Incrementar la percepción de riesgo
- Fomentar el fortalecimiento institucional los impuestos y contribuciones

Administra los siguientes Impuestos y contribuciones:

- Ingresos Brutos
- Inmobiliario
- Automotor
- Embarcaciones Deportivas
- Sellos
- Planes de Regularización de Deudas
- Contribución Provincial de Energía
- Contribución Fondo Provincial de la Vivienda
- Contribución Fondo Provincial Educación

La cantidad de registros por impuesto es de:

- Inmobiliario: 5.584.638
- IIBB: 1.652.567
- Automotores: 2.896.309
- Agentes de Recaudación: 32.812
- Embarcaciones: 53.294

Es de suma importancia incorporar tecnología que simplifique el pago de los impuestos y dinamice los trámites que realizan los contribuyentes, favoreciendo así el cumplimiento tributario.

En línea con todo lo anterior mencionado, la Agencia se encuentra en un proceso de consolidación de toda la información existente de los contribuyentes de la Provincia de Buenos Aires.

La carga de trabajo que demandan los objetivos operativos y estratégicos de la Agencia, las capacidades y herramientas necesarias para poder establecer un proceso de mejora continua en sobre los datos y la necesidad de contar, en corto tiempo, con una mejora en los atributos de los datos de los contribuyentes, hacen necesario contar con un servicio que permita lograr los objetivos esperados del presente proyecto en los tiempos acordados y con los recursos necesarios.

Dentro de la Agencia existe un gran volumen de datos de contribuyentes, que actualmente se encuentra desactualizado y con la necesidad de ser revisado, calificado, reestructurado y optimizado a través de un proceso de calidad de datos sobre todo el padrón de contribuyentes de cada uno de los impuestos.

Para cumplir con las misiones y mejorar la calidad del servicio, la Agencia se encuentra en un proceso de perfilar, sobre una base de datos relacional, normalizada, los datos del padrón de Contribuyentes donde se pueda aplicar técnicas de calidad de datos.

La oportunidad que se presenta es la migración de estructura de datos y de la plataforma que lo soporta.

El Core en el cual se apoya el negocio de la agencia es ADABAS – Natural y se requiere comenzar a migrar la lógica de negocio y la información a una Base de Datos sobre tecnología ORACLE, dejando de trabajar sobre modelos jerárquicos y pasando a construir sobre modelos relacionados y normalizaos.

En la primera etapa del proceso se realizó un análisis exhaustivo de cada file que compone este modelo y más en detalle, también se analizó cada campo que forman este conjunto de files para poder desarrollar los controles de calidad y las reglas de remediación de los errores detectados.

Esto permitirá que tanto la extracción de información de estos ambientes se realice de manera automática, transparente y auditable, protegiendo la información que se transporta y aplicando procesos de transformación y calidad de datos.

Este modelo dará sustento a los procesos críticos del negocio de ARBA, como, por ejemplo, al proceso de emisiones, el cual utiliza la información de

los domicilios, calles y demás datos territoriales para poder entregar la boleta en tiempo y forma.

3 Solución

Data Governance (DG) o Gobierno de Datos⁶, es el manejo total de la responsabilidad, disponibilidad, usabilidad, integridad y seguridad de los datos de una empresa. Un programa de gobierno de datos se conforma, entre otras cuestiones, de un consejo que define las políticas y un conjunto de procedimientos y sus correspondientes planes de ejecución.

Aplicar un proceso de calidad de datos, permite elevar el nivel de exactitud, completitud, integridad, actualización, coherencia, relevancia y confiabilidad de la información, permitiendo identificar y corregir problemas de:

- Completitud
- Reglas de Negocio
- Coherencia
- Integridad paramétrica
- Integridad referencial

Para llevar adelante un proceso de calidad de datos es necesario ejecutar las siguientes fases:

- Analizar los factores que afectan la calidad y el uso de los datos.
- Definir un enfoque sistemático para evaluar los niveles de calidad de los datos.
- Por último, se corrigen los errores de datos detectados y elimina las causas raíz de errores de datos

Siendo más específicos, los pasos que se realizan, son los que se mencionan a continuación, y se retroalimentan para llevar a un estado óptimo la calidad de los datos:

1. **Definición:** Definir términos de negocio, fuentes de datos y requerimientos de negocio.
2. **Perfilar:** Analizar los datos para comprender cómo están constituidos, sus fuentes y estado de calidad actual.
3. **Diseñar:** Construir los flujos de trabajo para fortalecer las reglas de negocio y construir repositorios consolidados.
4. **Resolución:** Diseñar los controles de QA para el monitoreo y resolución automática.
5. **Evaluar:** Comparar los resultados esperados contra los reales para poder comunicar efectivamente.

⁶ Data Governance - https://en.wikipedia.org/wiki/Data_governance

6. **Controlar:** Mejorar los procesos y sistemas a través del plan de remediación de datos.

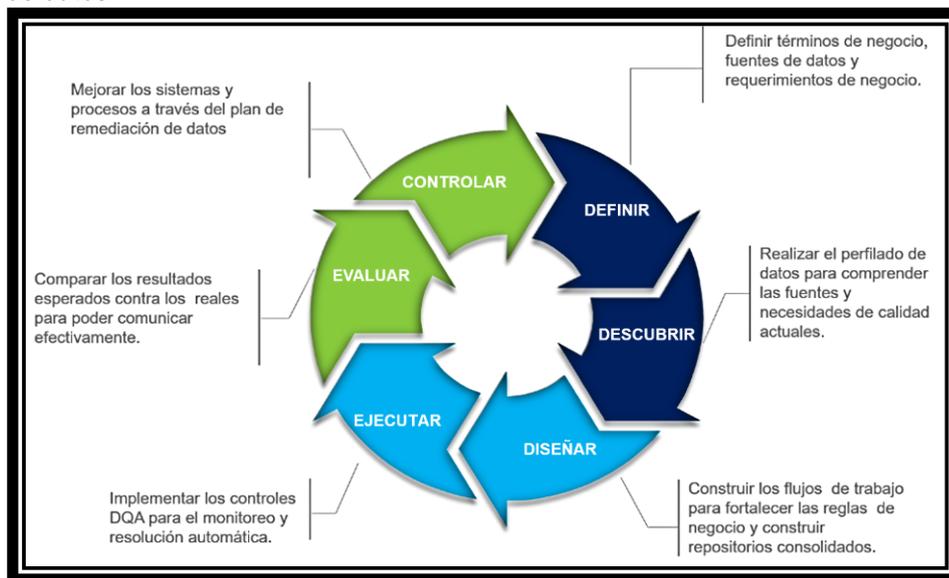


Fig. 1

3.1 Descripción del proyecto

La Solución consiste de un proceso de normalización de la información que se divide en varias etapas que se mencionan a continuación:

- Perfilado de Datos
- Limpieza de Datos
- Mejora de Calidad de Datos
- Fusión de Datos Repetidos
- Generación de Documentación (Informes/Reportes)

La tarea de Perfilado de Datos deberá:

- Confeccionar un diccionario de términos y aplicaciones existentes
- Identificar la información crítica y relevante de los contribuyentes
- Identificar las distintas fuentes de datos, internas y externas
- Realizar un Informe detallado de los problemas encontrados en primera instancia y recomendaciones para su saneamiento, que guiarán el trabajo futuro (siguientes etapas del proceso de Calidad de Datos).
- Determinar métodos de producción de la información de cada fuente (automática/manual) y Changed Data Capture (CDC).

- Establecer métodos de calidad a aplicar en cada fuente (Correspondencia, De duplicación, Normalización, etc.)
- Determinar quiénes son los "dueños" de los activos de Información. Es decir, a través de qué aplicaciones se generan los distintos grupos de datos, determinar qué programas/aplicativos son los generadores y en caso de ser posible, especificar qué grupos de Departamentos son los que utilizan dichos aplicativos.
- Determinar quiénes son los dueños y quienes son lo que actualizan los datos.

Las tareas que se realizarán para llevar adelante el Data Cleansing son las siguientes:

- Establecer los métodos para realizar el Data Cleansing, con sus dependencias técnicas y/o funcionales
- Aplicar las técnicas requeridas según el método de calidad, la complejidad de procesamiento y las capacidades de flujo.
- Establecer una metodología para la limpieza de los datos.
- Documentar recomendaciones que permitan limpiar los datos progresivamente en las fuentes de datos.
- Documentar buenas prácticas y modificaciones necesarias para evitar seguir ensuciando los datos ya procesados.

Se requieren las siguientes tareas para llevar adelante la etapa de Quality Data:

- Análisis de datos faltantes en nuestras bases
- Análisis de datos mal cargados o erróneos
- Análisis de factibilidad de la posibilidad de mejora de los datos detectados en los puntos previos.
- Establecer una metodología para realizar una actualización de los datos detectados en los puntos, en caso que esta tarea sea factible.
- Documentar todas las tareas realizadas en esta etapa, detallando los registros (filas, columnas, etc.) encontrados como faltantes, y pasos que se siguieron para mejorar estos datos.

Se requieren las siguientes tareas para llevar adelante la etapa de Data Matching:

A través de este proceso se busca detectar todo tipo de inconsistencias en la BB.DD.

- Detección de datos duplicados
- Relación entre dos fuentes de datos que no tienen campos de unión
- Detección de unidades familiares (householding)

Estrategia:

El Proyecto se llevó adelante definiendo un equipo de proyecto y utilizando la metodología implementada por GGTI, basada en PMI, para la gestión de proyectos tecnológicos.

Como todo proyecto, se detectaron a las áreas interesadas, se definió un cronograma, un líder de proyecto, se definió el alcance y se comenzó la tarea a partir de una reunión de kick-off.

Como parte de la estrategia se le dio principal importancia en la detección de los distintos data owner's y también a su convencimiento en cuanto a mostrarles cuán importante es tener una foto real del estado de la información en la empresa y también convencerlos para que entiendan lo importante que de su participación en este proyecto.

Los data owners tienen un papel importante en la definición del proceso de gobierno de datos, sin ellos no es posible tomar decisiones sobre la transformación y definición de las reglas de remediación automáticas sobre los datos.

Fue indispensable hacer detección de cada área interesada, en este tipo de proyectos es muy importante no olvidar a ninguno, ya sea beneficiados o no el proyecto.

Se focalizo cada etapa del proyecto en los hitos definidos en el acta de constitución y Gantt del proyecto, evaluando cada cambio de alcance de manera minuciosa, para evitar desviar el foco en los objetivos definidos para el proyecto.

4 Innovación e Inédito

La herramienta que se utiliza para llevar adelante el proceso es SAS Data Connector, es la primera implementación de este tipo de conector sobre mainframe con z/OS y ADABAS. Esto, no solo permite aplicar todos los pasos para aplicar calidad, sino que también posee un conector para realizar las extracciones de los files indicados y perfilarlos hacia un motor Oracle destinado para tal sentido.

Es la primera vez en la historia de la agencia que se avanza sobre un proyecto con estas características con la intención de mejorar la calidad de y precisión de los servicios y establecer un proceso que permita gobernar los datos de los contribuyentes.

Si bien se contrató una empresa con experiencia en la materia, la complejidad del negocio y de la tecnología utilizada en arba hace que no se puede hacer Benchmarking con los trabajos realizados en otras compañías,

5 Beneficiarios

El proyecto de gobierno de datos implica directamente revisar, actualizar, calificar, reestructurar y optimizar la información de los contribuyentes, pero también presenta una oportunidad de migración de tecnología que facilite el desarrollo, mantenimiento de aplicaciones como parte de un plan de innovación que se alinea directamente con los propósitos tanto de ARBA.

Este proceso es transparente para el contribuyente y lo beneficia directamente permitiéndole tener acceso información más precisa, consistente, actualizada y en línea de sus datos tributarios como de su historial como contribuyente de la provincia de buenos aires.

6 Relevancia para el Interés Público

Establecer políticas y procedimientos para la gobernanza de datos en la agencia, en conjunto con un proyecto de calidad de datos, permite mejorar la calidad del servicio que proveen a sus clientes, a partir del uso de información precisa y completa, permitiendo, además, ser eficientes en el manejo de recursos humanos, tecnológicos y económicos para llevar adelante sus procesos productivos y de servicios.

7 Viabilidad Técnica, Financiera y Política Organizacional

Es importante que todo proyecto esté alineado con los objetivos estratégicos de la organización y deberá tener el apoyo necesario para poder definir responsabilidades, flujos de trabajo y controles necesarios sobre los datos que estén correctamente validados con los dueños de la información.

La agencia puso en marcha este proyecto que se planteó con un plazo de 8 meses de trabajo donde se contrató a una empresa con experiencia en el mercado sobre calidad de datos.

La coordinación, seguimiento, control de avance del proyecto es llevada a delante por el dpto. de administración de proyectos de la gerencia de Gestión de Proyecto y servicios de GGTI.

Para el avance del proyecto las áreas de negocio deberán prestar total disponibilidad para proveer el conocimiento, ya sea de negocio como la descripción funcional que tienen las aplicaciones que están desarrolladas, para llevar adelante las etapas que involucran la calidad sobre el padrón de contribuyentes.

Esto permitirá consolidar la información de todos los contribuyentes de la Provincia de Buenos Aires en una tecnología más amigable con la nueva generación de aplicaciones que tiene en sus planes la agencia como parte de un proceso de modernización y mejora en la calidad de sus servicios.

8 Facilidad de Reproducción

El proyecto debe tener el aval y apoyo correspondiente de todas las partes involucradas durante la planificación y ejecución del mismo.

Llevar un proyecto con estas características involucra consolidar un equipo de profesionales en la materia. Es importante conocer y establecer el conjunto de datos sobre los cuales se hará el proceso, definiendo el alcance, identificando claramente a cada uno de los interesados ya sean "owner" de los datos como responsables del negocio que hace uso de los mismos. Definir los dueños de los datos no solo es importante para el gobierno de los datos, sino que define además responsabilidades y funciones claras sobre cada uno de los datos de la organización. Las etapas que se desarrollaron están basadas en procesos estándares de calidad de datos y el proceso definido resultante es independiente de la infraestructura tecnológica con la que cuentan los organismos del estado.

9 Ambiente de Hardware y Software

La solución consiste en la incorporación de la herramienta SAS DataFlux Data Management Studio 2.7, la cual interactúa entre las fuentes de datos existentes y el nuevo Modelo de Datos construido con el objetivo de ser consumido a futuro por las nuevas aplicaciones a ser implementadas por ARBA.

Para que dicha herramienta pueda conectarse y extraer datos desde fuente ADABAS, se implementó un servidor Linux que, mediante un conector SAS, tiene la capacidad de recibir los datos del servidor correspondiente a ADABAS (en lenguaje Natural) y realizar todo el procesamiento de Calidad requerido.

La herramienta mencionada aplica las reglas de Calidad de Datos identificando los problemas y remediándolos en origen, alimentando a su vez el Nuevo Modelo de Datos construido en Tercera Forma Normal.

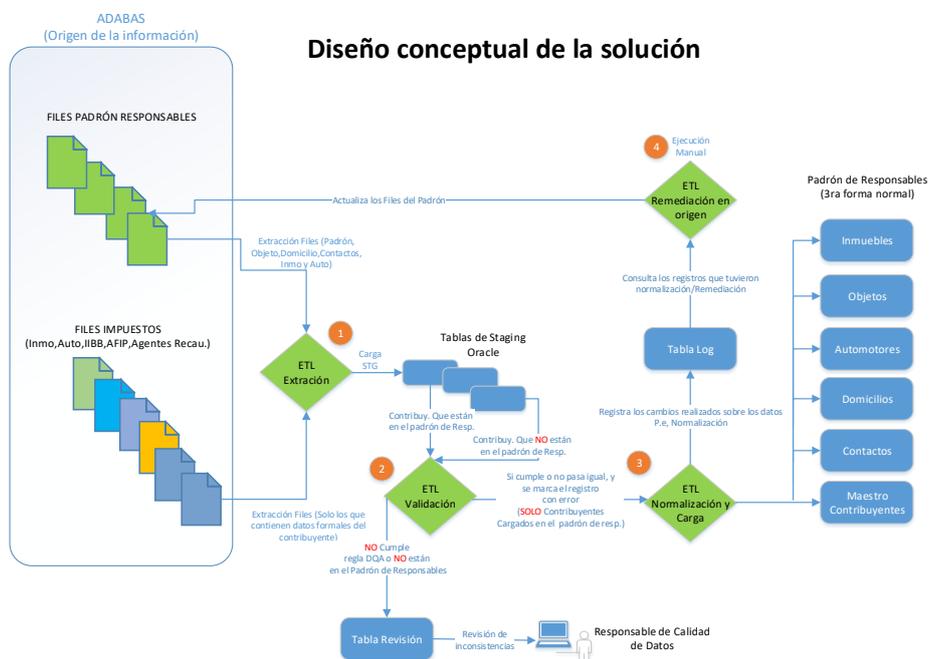


Fig. 2

10 Desafío

Como todo Proyecto, es importante cumplir con los tiempos pautados para lograr llevar adelante todo el proceso de calidad de datos sobre el padrón de contribuyentes y que el producto resultado permita obtener una base de datos de referencia donde la información cumpla con las medidas de calidad necesarias para poder proveerle al contribuyente la información en tiempo y forma.

El principal objetivo del proceso de Gobierno de Datos es asegurar la homogeneidad y consistencia de los datos desde la fuentes de origen(ADABAS) hacia el repositorio final de explotación de la información(ORACLE) y fundamentalmente dejar establecido un proceso que, sustente políticas, responsabilidades, tecnología y procedimientos claros sobre la administración de la información, donde también defina y tome decisiones sobre los datos de los contribuyentes de la Provincia de Buenos Aires ⁷

⁷ <https://ciclusgroup.wordpress.com/2013/10/02/la-importancia-de-realizar-un-analisis-de-calidad-de-datos/> - "La Importancia de realizar un Análisis de Calidad de Datos" - José Casachahua

Como principales objetivos del proyecto es esperable:

- Establecer un proceso de calidad de datos que asegure una mejora continua en las tareas de recuperación, análisis, corrección y actualización de los datos.
- Identificar las fuentes de información que contengan los datos que necesitan ser corregidos y actualizados.
- Identificar el origen de los datos con baja calidad para poder corregir y asegurar su creación y guarda.
- Crear un modelo de base de datos normalizado en ORACLE que permita definir el nuevo modelo transaccional del padrón de contribuyentes.
- Elevar el nivel de calidad de la información de la información del padrón de contribuyentes.
- Consolidar una Base de datos Transaccional que además cumpla con las características que establece como resultado un proceso de calidad de datos (precisión, oportunidad, relevancia y comprensibilidad).

11 Referencias Bibliográficas

1. Deloitte named global leader in Business Analytics Services by Gartner – Deloitte. <https://www2.deloitte.com/global/en/pages/about-deloitte/articles/deloitte-named-global-leader-business-analytics-services.html>.
2. Enabling Essential Data Governance for Successful Big Data Architecture Deployment – Gartner. <https://www.gartner.com/doc/3842163?ref=SiteSearch&sthkw=data%20governance&fnl=search&srcl=1-3478922254>
3. "The DGI Data Governance Framework" - Gwen Thomas, The Data Governance Institute. http://www.datagovernance.com/wp-content/uploads/2014/11/dgi_framework.pdf
4. Hopwood, Peter (June 2008). "Data Governance: One Size Does Not Fit All". DM Review Magazine. Archived from the original on 2008-10-02.
5. SAS DATA GOVERNANCE - https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/sas-data-governance-framework-107325.pdf
6. "La Agencia de Recaudación de la Provincia de Buenos Aires hacia un Gobierno Cognitivo" – Mg. Sandra D'Agostino
7. "Calidad de Datos" – Information Builders <http://www.informationbuilders.es/data-quality>

8. "El valor de la gestión de datos"- powerdata
<https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/368784/introduccion-a-la-calidad-de-datos-definicion-control-y-beneficios>
9. "La Importancia de realizar un Análisis de Calidad de Datos" - José Casachahua
<https://ciclusgroup.wordpress.com/2013/10/02/la-importancia-de-realizar-un-analisis-de-calidad-de-datos/>