

Fundación Nacional para el Desarrollo
Centro de Monitoreo e Incidencia Fiscal

DISEÑO DE LA PLATAFORMA DE DATOS

Versión final - Marzo 2020

Contenido

1 Introducción.....	3
2 Identificación de buenas prácticas.....	4
3 Necesidades de usuarios.....	7
4 Requerimientos funcionales.....	9
5 Diseño conceptual.....	9
6 Diseño lógico.....	11
7 Diseño de procesos.....	13
8 Monitoreo y evaluación.....	17
9 Referencias.....	18

1 Introducción

La Fundación Nacional para el Desarrollo está ejecutando el proyecto “Monitoreo ciudadano de la transparencia fiscal en El Salvador”. Este contribuye a la transparencia fiscal y la lucha contra la corrupción a través del uso intensivo de bases de datos y de aplicaciones informáticas. Consiste en un observatorio ciudadano que da seguimiento al presupuesto general del Estado, orientado hacia la promoción del acceso, uso y análisis de la información financiera, en formatos que sean asequibles, de fácil acceso y comprensibles para la población en general.

En ese marco, es necesario desarrollar la plataforma de datos del observatorio (el cual ha sido denominado “Centro de Monitoreo e Incidencia Fiscal”). La plataforma sirve para reconstruir procesos y transacciones, encontrar patrones o identificar anomalías que apuntan hacia hallazgos de corrupción o de uso ineficiente de los recursos públicos. El uso intensivo e integrado de bases de datos permite acceder a conjuntos detallados de información, como gastos reservados, viajes, medicamentos, obras de infraestructura, compras públicas y otros. Esta herramienta también permitirá el monitoreo del destino de los proyectos financiados por la cooperación internacional.

Este documento contiene el diseño de la plataforma de datos. Se ha hecho una revisión de literatura y de plataformas de finanzas públicas de otros países, en particular las operadas por organizaciones civiles. Los usuarios proyectados de la plataforma han sido agrupados en tres categorías: especialistas, interesados directos y público general. Por cada categoría se han listado requerimientos particulares en cuanto al contenido y formato de la información. Con base en buenas prácticas sugeridas por la literatura, ejemplos de otras plataformas y las necesidades de los usuarios, se han definido los requerimientos sobre las funciones que debe desempeñar la plataforma. Por último, se presenta el diseño conceptual, el lógico y el de procesos. El primero consiste en un diagrama de bloques que representa los componentes de la plataforma y la función que tendrá cada uno de ellos. El diseño lógico es la descripción de los estados y de las entradas y salidas de cada componente. El diseño de procesos es la descripción de los pasos que se siguen en cada proceso para transformar las entradas en salidas. Luego contiene una sección de monitoreo y evaluación.

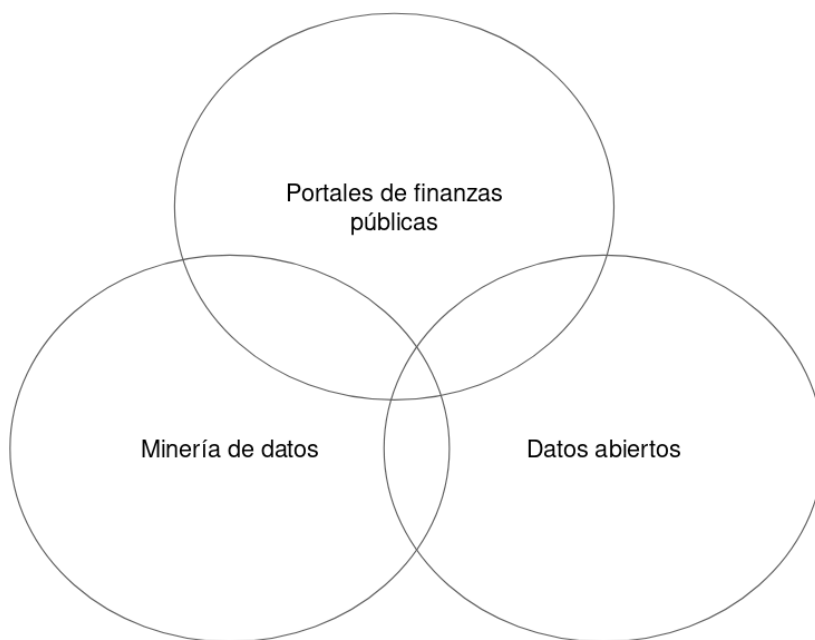
El modelo de desarrollo de la plataforma ha sido en espiral. Es decir, con los elementos iniciales de diseño se ha procedido a la implementación de los primeros componentes de la misma. En los siguientes ciclos, el diseño ha sido complementado con los aspectos lógicos y de procesos de la plataforma, así como con los instrumentos de monitoreo y evaluación del desempeño de la misma. Luego de eso se han hecho ciclos de mejora hasta contar con un producto que satisfaga los objetivos del proyecto y las necesidades de los usuarios.

2 Identificación de buenas prácticas

En esta sección se citan estándares e implementaciones de plataformas de datos como ejemplos de buenas prácticas. Los estándares son guías sugeridas por diferentes autores u organismos mediante las cuales se dan orientaciones generales para el diseño e implementación de plataformas de datos. Por lo general, los estándares surgen del acuerdo entre múltiples partes interesadas o a partir de lecciones identificadas en la evaluación de experiencias concretas. En cuanto a la implementación de plataformas de datos se ha seleccionado, de forma intencionada, una muestra de portales de datos sobre finanzas públicas. La intención es contar con referencias de partida para luego, en siguientes avances, ampliar la exploración en forma comparativa. En esta revisión de buenas prácticas es importante tener presente que el campo de las plataformas de datos sobre finanzas públicas está abierto a la innovación. Por ello, los estándares y ejemplos de implementaciones deben ser tomadas como pautas para la experimentación y no como recetas o prescripciones a seguir.

Figura 1.

Conceptos claves de buenas prácticas



La exposición de las buenas prácticas ha sido organizada alrededor de tres conceptos: minería de datos, datos abiertos y portales sobre finanzas públicas (ver Figura 1). Los primeros dos tópicos se abordan en forma general, sin una referencia directa a la cuestión de las finanzas públicas. La minería

de datos es un proceso para extraer información útil, comprensible y novedosa del procesamiento de grandes volúmenes de datos (Moine, Haedo y Gordillo, 2011). Los datos por lo general provienen del registro de transacciones cotidianas y se analizan no mediante muestras estadísticas, sino sobre el universo completo de datos disponibles. Los datos abiertos, o en forma más precisa la apertura de los datos, consiste en un proceso para poner a disposición del público datos procedentes de diferentes organizaciones en un alto nivel de detalle, en formatos de fácil procesamiento y con licencias libres. Poner los datos a disposición de la sociedad permite que cualquier persona pueda construir con ellos un nuevo producto que resulta en datos de mayor valor, en conocimientos o en servicios (Concha y Naser, 2012). Por último, los portales sobre finanzas públicas no son otra cosa que sitios web dedicados a la publicación de información sobre las finanzas públicas como una herramienta para mejorar la transparencia y la rendición de cuentas (Solana, 2004). Tradicionalmente estos portales han sido mantenidos por los gobiernos, en particular los ministerios de finanzas, pero más recientemente existen también portales operados por centros de investigación y organizaciones de sociedad civil.

Las metodologías estandarizadas de minería de datos se diferencian en principio por el número de pasos que comprenden. Esto responde al nivel de atención que se le presta a aspectos como la naturaleza del problema bajo estudio, el contexto de trabajo, las fuentes y calidad de los datos, la dinámica de análisis de los datos y la canalización de la información extraída. Shafique y Qaiser (2014) hacen una comparación de tres de las metodologías más conocidas: KDD, CRISP-DM y SEMMA. KDD (Knowledge Discovery Databases) requiere conocimiento previo y una comprensión relevante del dominio de aplicación y de los objetivos del proyecto de datos. CRISP-DM (CRoss-Industry Standard Process for Data Mining) pretende establecer un marco y guías uniformes para practicantes de minería de datos. SEMMA (Sample, Explore, Modify, Model, and Access) procura el entendimiento, la organización, el desarrollo y el mantenimiento de los proyectos de datos.

Cuadro 1.

Metodologías de minería de datos

KDD	CRISP-DM	SEMMA
1. Desarrollo y entendimiento del dominio de aplicación 2. Creación del conjunto de datos objetivo 3. Limpieza y preprocesamiento de datos 4. Transformación de datos 5. Selección de la tarea adecuada de minería de datos 6. Selección del algoritmo de minería de	1. Comprensión del negocio 2. Comprensión de los datos 3. Preparación de datos 4. Modelado 5. Evaluación de modelos 6. Implementación	1. Muestreo 2. Exploración 3. Modificación 4. Modelado 5. Acceso

datos adecuado 7. Empleo del algoritmo de minería de datos 8. Interpretación de patrones encontrados en los datos 9. Uso del conocimiento descubierto		
--	--	--

Fuente: elaboración propia con base en Shafique y Qaiser (2014).

En el Cuadro 1 se presenta una comparación de los pasos sugeridos por las metodologías de minería de datos citadas. KDD es la más detallada, poniendo énfasis tanto en la comprensión del problema como en el proceso para encontrar una solución. CRISP-DM considera el modelo en forma más general y asume que la solución será utilizada luego como parte de un sistema más grande; de ahí que concluye en la implementación de dicha solución. Por su parte, SEMMA está más orientada a la parte operativa del proceso de minería de datos.

Un conjunto de datos está abierto si cualquier persona es libre de acceder a él, usarlo y distribuirlo. Según Kitchin (2014), el carácter de abierto de los datos es en relación con el hecho de que estén disponibles de manera completa, en forma modificable y con costos de reproducción razonables. Además, la licencia no debe ser restrictiva ni discriminatoria; es decir que debe permitirle a cualquier persona hacer modificaciones y trabajos derivados, así como redistribuir los datos. Kitchin también cita las características de los datos abiertos aplicables a los gobiernos; en ese sentido, los datos deben ser completos, de fuente primaria, oportunos, accesibles, procesables, no discriminatorios, no propietarios, con licencias libres y verificables.

Cuadro 2

Ejemplos de plataformas de datos sobre finanzas públicas

Portal	Responsables	Descripción
WikiBudgets https://www.wikibudgets.org	Daniel Sedlacek y Wojciech Piekutowski	Visualizaciones de presupuestos de instituciones públicas alrededor del mundo.
Africa Debt Monitor https://www.cabri-sbo.org/en/budgets-in-africa/africa-debt-monitor	The Collaborative Africa Budget Reform Initiative (CABRI)	Información compartida por los gobiernos de África sobre las políticas, operaciones y saldos de la deuda con el propósito de evaluar la sostenibilidad del endeudamiento público.
RindeCuentas.org https://www.rindecuentas.org	Proyecto sobre Organización, Desarrollo, Educación e Investigación (PODER, México)	Fomentar la rendición de cuentas empresarial mostrando el fenómeno de la captura del estado a través de las

		historias periodísticas.
Privilegios Fiscales https://privilegiosfiscales.fundar.org.mx	Fundar, Centro de análisis e investigación (México)	Sitio consulta de una base de datos sobre privilegios fiscales concedidos por el gobierno federal. La base de datos se obtuvo por medio de un litigio de acceso a la información.
Monitor Presupuestario http://monitorpresupuestario.acij.org.ar/monitor	Asociación Civil por la Igualdad y la Justicia (ACIJ, Argentina)	Información presupuestaria de asignación y ejecución de diferentes organismos públicos

Fuente: elaboración propia

Según Solana (2004), los portales sobre finanzas públicas ofrecen tres tipos de información: (a) no transaccional, por ejemplo leyes, manuales y glosarios; (b) resúmenes de transacciones, como estados financieros y presupuestarios y saldos de la deuda; e (c) información transaccional, por ejemplo la ejecución presupuestaria, los registros contables y las operaciones de compras públicas. Con base en la revisión de lecciones aprendidas, Dener & Young (2013) sugieren los siguientes principios para los portales de datos presupuestarios: oportunidad y comprensibilidad de la información, revelación de detalles acerca de los sistemas de información subyacentes, disponibilidad de consultas y reportes definidos por el usuario, confiabilidad e interoperabilidad de los datos, autenticidad de las fuentes de datos, calidad de la presentación y promoción para el uso efectivo de los datos. Estos autores sostienen que la cuestión clave es poder ver hacia dónde va el dinero público. Como ilustración, en el cuadro 2 se citan ejemplos de portales y plataformas de datos operadas por organizaciones civiles, personas y entidades internacionales.

3 Necesidades de usuarios

En esta sección se agrupan en categorías los usuarios que se considera que tienen interés en la plataforma de datos y a la vez se anticipan las necesidades de información de estos. Esta información tiene que ser validada en la medida en que va entrando en operación la plataforma de datos, por medio de los instrumentos de monitoreo y evaluación de la misma. Las necesidades de información si bien pueden responder a un patrón, se debe asumir que son cambiantes y por ello la plataforma de datos es concebida de tal forma que se pueda ir adaptando con el tiempo.

Las categorías de usuarios definidas son especialistas, interesados directos y público general. Los especialistas (ESP) son personas con conocimiento detallado de las finanzas públicas y que por razón de su ejercicio profesional interactúan frecuentemente con datos fiscales y presupuestarios. Los interesados directos (ID) son personas que se desempeñan profesionalmente en ámbitos distintos de

las finanzas públicas pero que requieren de datos fiscales o presupuestarios para comprender de mejor manera sus propios contextos y problemáticas. El público general (PG) son personas que ocasionalmente tienen la necesidad de consultar un dato fiscal o presupuestario para responder a preguntas de coyuntura o de interés particular. Estas categorías de usuarios van de la más particular a la más general. Es decir que, para efectos prácticos, se asume que dentro del público general están incluidos los interesados directos y que dentro de los interesados directos están incluidos los especialistas.

El Cuadro 3 presenta diferentes tipos de información y niveles de acceso a la misma identificando qué usuarios pueden tener mayor interés en cada intersección. Los tipos de información han sido definidos principalmente con un lógica temporal que inicia en los momentos de formulación y concluye en los momentos de liquidación y revisión de cuentas, como lo son por ejemplo el ciclo presupuestario y el ciclo de endeudamiento público.

Cuadro 3

Necesidades de información

Categorías de información	Nivel de acceso		
	Visualización	Consulta	Datasets
Variables macroeconómicas		ESP	
Proyecciones de mediano plazo		ESP	
Proyecciones preliminares	PG	ID	ESP
Proyectos legislativos	PG	ID	ESP
Instrumentos votados (presupuestos, deudas, impuestos, etc.)	PG	PG	ID
Datos de ejecución	PG	ID	ESP
Informes de resultados	PG	PG	ID
Auditorías y evaluaciones	PG	ID	ESP

Fuente: elaboración propia.

4 Requerimientos funcionales

Con base en lo descrito en las secciones anteriores, en este apartado se definen los requerimientos funcionales de la plataforma de datos. Estos requerimientos son los que gobiernan el desarrollo de la plataforma. Cada uno de ellos irá alcanzando nivel de especificidad en la medida que se construyan los diferentes componentes. Tal especificidad será explicitada y documentada debidamente.

1. Metodología de minería de datos basada en KDD
2. Datos disponibles en formato procesable (por ejemplo CSV, JSON o similares)
3. Licencia de uso libre, únicamente con restricción de atribución (reconocimiento de la fuente)
4. Sitio con acceso libre (sin requerimiento de registro)
5. Uso de software con licencia libre o de código abierto
6. Publicación de código fuente
7. Diseño diferenciado para usuarios especialistas, interesados directos y público general
8. Aplicación de políticas de privacidad para la protección de los datos de los usuarios
9. Incluir en orden de prioridad información transaccional, resúmenes e información no transaccional
10. Ofrecer archivos de metadatos para informar sobre la fuente de los datos y la interpretación de los mismos
11. Tres niveles de acceso: visualización, consulta y descarga de datasets
12. Organizar la información conforme las etapas de los ciclos fiscales y presupuestarios
13. Contar con un calendario de publicación de información
14. Proporcionar ayuda y herramientas de promoción, para atraer a usuarios y facilitar el uso de la plataforma de datos

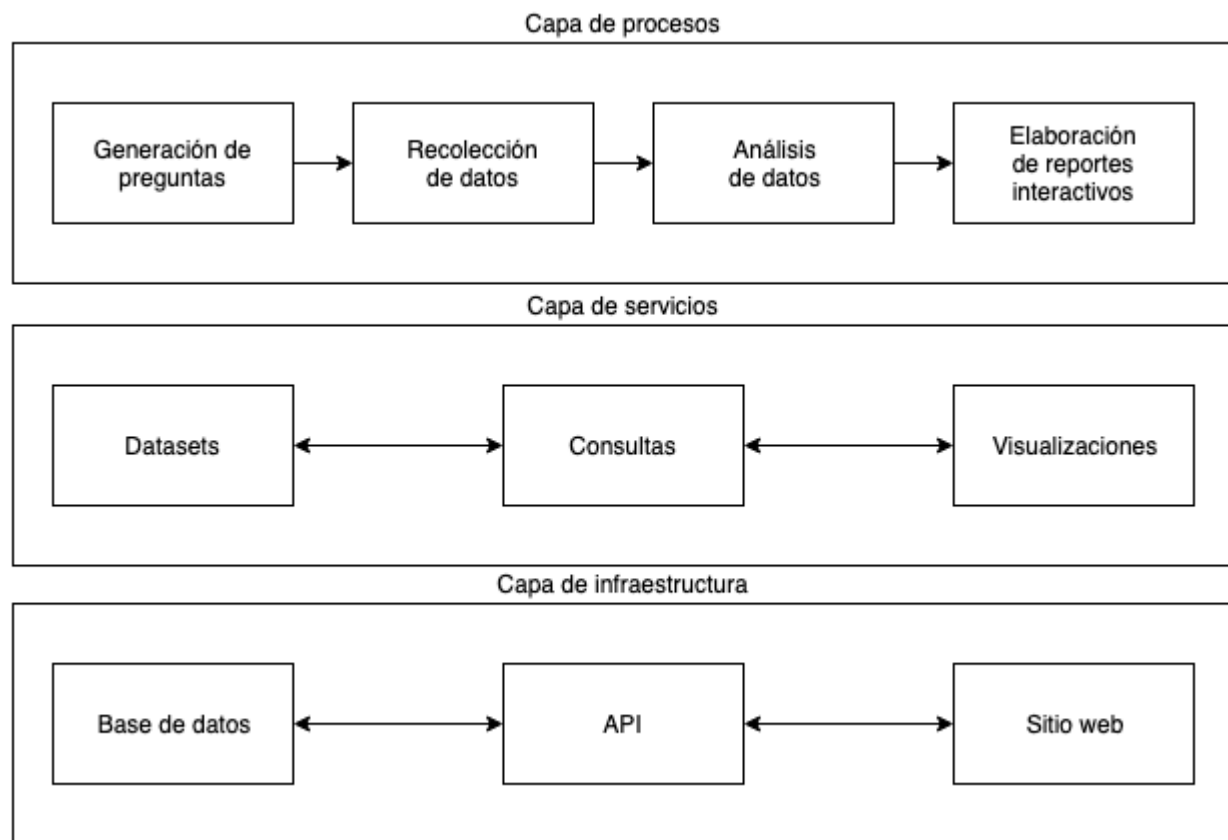
5 Diseño conceptual

En la Figura 2 se presenta el diseño conceptual de la plataforma de datos bajo un modelo de capas. La capa superior es la de procesos, definida en el dominio del problema. En esta, la metodología KDD ha sido condensada en un proceso de cuatro pasos. La capa intermedia es la de los servicios de la plataforma, es decir los mecanismos permanentes mediante los cuales los usuarios interactúan,

independiente del tópico de interés. La capa inferior es la de infraestructura, es decir el soporte tecnológico para la operación de la plataforma.

Figura 2

Diseño conceptual de la plataforma de datos



La capa de procesos inicia con la generación de preguntas de indagación. Estas surgen de la problematización de la situación de las finanzas públicas en relación con el contexto social (presente o futuro). Este es el factor que le imprime carácter dinámico y actualidad a la plataforma. Para responder a las preguntas de indagación es necesario identificar los datos a partir de los cuales se puedan construir respuestas. En este caso, los datos generalmente son datasets sobre diferentes componentes de las finanzas públicas como impuestos, presupuestos, deudas o compras, entre otras posibilidades, los cuales se relacionan con indicadores o estadísticas del entorno. El análisis de los datos puede incluir en las etapas preliminares operaciones rutinarias de sumarización y de cruce de datos, pero a medida que se avance con la plataforma la idea es generar modelos explicativos o predictivos en los asuntos

bajo indagación. La información resultante de los análisis se divulgan por defecto por medio de reportes interactivos (dashboards) en los cuales los usuarios puedan ajustar parámetros y obtener diferentes vistas, así como tener acceso a los datasets.

Como se mencionó, la capa de servicios es la que contiene los mecanismos permanentes mediante los cuales los usuarios interactúan con la plataforma. Para tal efecto, se han concebido tres mecanismos de acceso: visualizaciones, consultas y datasets. Las visualizaciones son representaciones gráficas de los datos, generalmente de carácter estadístico, como distribuciones y series temporales. Las consultas son controles por medio de los cuales los usuarios pueden ajustar los parámetros para la selección de datos de interés, por ejemplo períodos, instituciones o rubros. Finalmente los datasets son la opción de acceso directo a los datos para descarga o transmisión. Hay que observar que tanto los datasets como las visualizaciones son adaptadas conforme los parámetros establecidos por los usuarios mediante las consultas.

La capa de infraestructura es el área detrás de bastidores de la plataforma. Tiene tres componentes, la base de datos, la API (application programming interface) y el sitio web, los cuales pueden estar distribuidos (ubicados en diferentes localidades). La base de datos es el conjunto de tablas en las que se van guardando en forma relacional los datos de la plataforma, de tal forma que se facilite la búsqueda y recuperación de los mismos. El sitio web es el conjunto de servicios para generar páginas web, en particular páginas dinámicas basadas en parámetros definidos por los usuarios. La API permite hacer el enlace entre la base de datos y el sitio web, y además es utilizada por servicios de terceros, lo que le da el carácter de abierto a la plataforma.

6 Diseño lógico

En esta sección se presenta la desagregación lógica de los componentes de la plataforma. Cada componente tiene asociados estados que definen la etapa en que este se encuentra en cuanto al procesamiento de información. Con base en los estados se define el flujo de trabajo de la plataforma en una estructura modular e interactiva. Los estados por lo general son secuenciales, lo que no excluye la posibilidad de que existan saltos o reiteraciones. Además, para cada componente se han definido sus entradas y salidas.

Cuadro 4

Diseño lógico de la capa de procesos

Componente	Estados	Entradas	Salidas
Generación de preguntas	1. Consulta 2. Discusión	- Comentarios de usuarios - Opinión de expertos	- Preguntas de indagación

	3. Formulación 4. Validación	- Noticias publicadas en medios de comunicación - Preguntas tentativas sugeridas en análisis previos	
Recolección de datos	1. Identificación de fuentes 2. Gestión de requerimientos 3. Preprocesamiento de datos	- Datos en línea - Datos proporcionados por medio de solicitudes de información - Datos recolectados ad-hoc	- Datasets estandarizados
Análisis de datos	1. Definición del procedimiento de análisis 2. Análisis descriptivo 3. Análisis inferencial* 4. Análisis predictivo* 5. Evaluación de modelos* * En aplicaciones avanzadas	- Datos pre-procesados	- Datos procesados - Tablas de datos - Visualizaciones preliminares
Elaboración de reportes interactivos	1. Esquema de contenido y presentación 2. Reporte en fase de desarrollo 3. Reporte en fase de producción 4. Reporte en fase de revisión	- Datos procesados - Tablas de datos - Visualizaciones preliminares - Citas de literatura	- Reporte formal - Reporte interactivo (dashboard)

Cuadro 5
Diseño lógico de la capa de servicios

Componente	Estados	Entradas	Salidas
Datasets	1. Datos crudos 2. Datos normalizados 3. Selecciones y vistas de datos	- Datos en línea - Datos proporcionados por medio de solicitudes de información - Datos recolectados ad-hoc	- Selecciones (SELECT) - Vistas (VIEW)
Consultas	1. Selección de dashboard 2. Filtros del dashboard (controles) 3. Controles de vistas (de	- Diccionarios de datos - Preferencias del usuario	- Parámetros de visualización y descarga de datos

Componente	Estados	Entradas	Salidas
	tablas y gráficas)		
Visualizaciones	1. Vistas por defecto 2. Vistas personalizadas	- Parámetros por defecto - Parámetros personalizados de visualización y descarga de datos	- Visualización gráfica - Visualización en tabla

Cuadro 6

Diseño lógico de la capa de infraestructura

Componente	Estados	Entradas	Salidas
Base de datos	1. Carga de datos 2. Consulta de datos (queries)	1. Datos normalizados 2. Diccionarios de datos	1. Consultas (queries)
API	1. Recepción de requerimiento 2. Preparación y envío de datos	1. Requerimientos REST (REpresentational State Transfer)	1. Catálogos (listas) 2. Detalles de entidades 3. Respuesta de actualización
Sitio web	1. Recepción de requerimiento HTTP(S) 2. Envío de respuesta	1. Requerimiento HTTP(s)	1. Contenidos web

7 Diseño de procesos

7.1 Estados de la capa de procesos

Generación de preguntas:

1. *Consulta:* Recurriendo a diferentes métodos de recolección de datos se hacen consultas con usuarios o expertos acerca de tópicos de interés. También se consultan noticias u opiniones publicadas en medios de comunicación social. Además, una vez que hay procesos concluidos dentro de la plataforma de datos, se revisan las preguntas tentativas que fueron formuladas en las secciones de discusión de los reportes respectivos.
2. *Discusión:* A nivel de equipo del proyecto o en la modalidad de consultas ampliadas, se revisan, organizan e integran los insumos obtenidos en las consultas. Se discuten posibles enfoques y se hacen consideraciones sobre el contexto y la coyuntura.

3. *Formulación*: Se elaboran preguntas tentativas, se valora su sentido, se ensayan en términos generales posibles respuestas, se priorizan y se afina su redacción.
4. *Validación*: Se valora la consistencia de las preguntas, la viabilidad de encontrar respuesta a las mismas, a la vez que se hacen consideraciones sobre los recursos disponibles para llevar adelante la indagación. Si es necesario, se hacen ajustes o acotaciones a las preguntas formuladas.

Recolección de datos:

1. *Identificación de fuentes*: Entre las fuentes de información se encuentran otros portales o servicios automáticos de datos, información que puede ser obtenida por medio de solicitudes de información o datos que requieren ser recolectados directamente.
2. *Gestión de requerimientos*: Se formulan los requerimientos de información y se envían o ejecutan. Se guarda registro de los eventos ocurridos durante la gestión de los requerimientos con la finalidad de determinar qué obstáculos han surgido que puedan ser indicativos de problemas de transparencia. Ciertos requerimientos de información se programan de manera periódica con la finalidad de ir actualizando los datos de la plataforma.
3. *Preprocesamiento de datos*: En el momento de recibir los datos, se revisan los archivos y se hacen ajustes para adecuar el formato, identificar inconsistencias o vacíos u otras tareas de preprocesamiento.

Análisis de datos:

1. *Definición del procedimiento de análisis*: Con base en los objetivos del análisis y la constatación de los datos obtenidos, se define el procedimiento de análisis y las herramientas que serán utilizadas. Esta definición puede tener lugar en forma interactiva, en la medida que se explora y se experimenta con los datos disponibles.
2. *Análisis descriptivo*: Consiste en el cálculo de estadísticas o métricas para describir los datos mediante operaciones de sumarización, filtrado, cruce de variables u otras.
3. *Análisis inferencial*: En algunos casos no se dispondrá del universo de datos, sino solo de muestras. En tales casos, si las muestras son válidas estadísticamente, se realizarán procedimientos de análisis para hacer generalizaciones a partir de dichas muestras de datos.
4. *Análisis predictivo*: Pueden incluir clasificaciones o regresiones mediante los cuales con base en ciertas observaciones se le atribuyen valores a variables no observadas.
5. *Evaluación de modelos*: Cuando se utilice más de un modelo de análisis para responder a determinadas preguntas, se hará una comparación entre los mismos para determinar las

ventajas y desventajas de cada uno de ellos o, en la medida de lo posible, elegir uno con base en las estimaciones de error.

Elaboración de reportes interactivos (dashboards):

1. *Esquema de contenido y presentación*: Consiste en un esquema de los tópicos que se presentarán en los reportes, la distribución de los contenidos en el medio de publicación (generalmente una página web) y el estilo de presentación. El esquema se puede ir definiendo en forma interactiva.
2. *Reporte en fase de desarrollo*: La elaboración de los reportes incluye la redacción de textos y la escritura de código. En fase de desarrollo, se implementa el esquema de contenido y presentación a través de procesos interactivos.
3. *Reporte en fase de producción*: En esta fase los reportes se ponen en línea para su consulta por parte de los usuarios. Siempre pueden tener lugar a cambios, pero estos se consideran menores y tiene solo el propósito de darle mantenimiento a los textos y al código.
4. *Reporte en fase de revisión*: Luego de cierto tiempo en que el reporte ha estado en fase de producción, pueden realizarle un actualización considerable del mismo o una reformulación. En tal caso se anuncia que estará en fase de revisión.

7.2 Estados de la capa de servicios

Datasets:

1. *Datos crudos*: Este estado se refiere a la forma cómo los datos son recibidos u obtenidos de las fuentes primarias o secundarias. Por lo general estos deberán ser modificados para su utilización, mediante operaciones de recodificación de variables, estandarización de atributos, definición de diccionarios y otras.
2. *Datos normalizados*: En este estado los datos ya se encuentran almacenados en la base de datos de la plataforma y cumplen con los estándares en cuanto a la denominación de los atributos y las relaciones entre estos. Los datos normalizados están listos para ser consultados, descargados o transmitidos. También se cuenta con metadatos para identificar las fuentes, describir sus atributos e incluir otras referencias útiles.
3. *Selecciones y vistas de datos*: Son las operaciones para la recuperación de los datos. Pueden incluir operaciones de sumarización, filtrado, cruce u otras. Los datos son puestos a disposición en formatos abiertos.

Consultas:

1. *Selección de dashboard*: Es un mecanismo de navegación para identificar y seleccionar el dashboard con el que desea interactuar el usuario. Este mecanismo puede incluir vistas temáticas o cronológicas, así como una herramienta de búsqueda. Este mecanismo será crítico en la medida en que el número de dashboards se incremente.
2. *Filtros del dashboard (controles)*: Los dashboards pueden contener diferentes controles, como entradas de texto, listas de selección, botones y otros, mediante los cuales el usuario puede personalizar la consulta.
3. *Controles de vistas (de tablas y gráficas)*: Las visualizaciones mediante tablas y gráficas también tienen sus propios controles para modificar la vista, por ejemplo para cambiar el orden de los datos o seleccionar los datos visualizados.

Visualizaciones:

1. *Vistas por defecto*: En general las visualizaciones por defectos se construyen con los parámetros más amplios, de tal forma que el usuario pueda luego ir acotando las vistas.
2. *Vistas personalizadas*: Mediante el uso de controles de selección de los datos y de la presentación de las vistas, los usuarios personalizan las visualizaciones.

7.3 Estados de la capa de infraestructura

Base de datos:

1. *Carga de datos*: La carga de datos se da por medio de operaciones INSERT y UPDATE (o sus equivalentes). Los datos serán almacenados en una base de datos que garantice la integridad de la información y que facilite la consulta de los datos.
2. *Consulta de datos (queries)*: La consulta de datos se da por medio de operaciones SELECT o FIND (o sus equivalente) para requerimientos inmediatos y por medio de operaciones VIEW para requerimientos más permanentes. En algunos casos, también se generan archivos secundarios con subconjuntos de datos de acceso frecuente.

API (application programming interface):

1. *Recepción de requerimiento*: Para la implementación de la interface se utiliza la metodología REST (REpresentational State Transfer) aplicada en direcciones web que permitan el acceso a los catálogos de datos (listados) y a las entidades (datos referidos a una observación, sujeto o asunto), por medio de operaciones GET, POST, UPDATE y DELETE (esta última de uso excepcional).

2. *Preparación y envío de datos*: Este estado conlleva la realización de las consultas a la base de datos, la preparación de los archivos y el envío de los mismos al servicio solicitante, que puede ser uno de los servicios de la plataforma o bien servicios de terceros.

Sitio web:

1. *Recepción de requerimiento HTTP(S)*: El sitio web es la interfaz de usuario por defecto de la plataforma. En el se publicarán los dashboards, documentos y otros recursos, a los cuales se podrá acceder mediante direcciones web permanentes.
2. *Envío de respuesta*: Al recibir el requerimiento de una página del sitio web, el servidor preparará los recursos necesarios para desplegar dicha página y los enviará al servicio solicitante.

8 Monitoreo y evaluación

En esta sección se definen los indicadores a utilizar para el monitoreo y la evaluación de la plataforma de datos. En el Cuadro 7 se han definido tres tipos de indicadores: de resultado, de proceso y de estado. El indicador de resultado corresponde a la evaluación, los de estado corresponden al monitoreo y los de proceso sirven para ambos propósitos, monitoreo y evaluación.

Cuadro 7

Indicadores de monitoreo y evaluación

Indicadores			Capas
De resultado	De proceso	De estado	
Los usuarios se sienten satisfechos con la información generada por la plataforma	Se han elaborado reportes que revelan patrones novedosos sobre la situación de las finanzas públicas en El Salvador	<ul style="list-style-type: none"> - Preguntas formuladas - Datos recolectados - Análisis de datos elaborados (memorias de cálculos) - Reportes publicados 	De procesos
	Se facilitado el acceso y la interacción de los usuarios con los datos de las finanzas públicas	<ul style="list-style-type: none"> - Visualizaciones interactivas habilitadas - Controles de consulta operando - Datasets disponibles 	De servicios
	Se han puesto a disposición herramientas tecnológicas innovadoras para la exploración e interacción con	<ul style="list-style-type: none"> - Integración de datasets fuente en la base de datos - API habilitada - Sitio web operando 	De infraestructura

Indicadores			Capas
De resultado	De proceso	De estado	
	los datos de las finanzas públicas		

El indicador global para la evaluación de los resultados se refiere a la satisfacción de los usuarios con la información generada por la plataforma de datos. En la plataforma se crearán las herramientas necesarias para, respetando las normas de privacidad, identificar los usuarios y obtener su retroalimentación acerca de los servicios ofrecidos. Además, se buscará realizar actividades de campo con ciertos usuarios para tener una evaluación directa de la iniciativa.

Los indicadores de proceso corresponden tanto al monitoreo como a la evaluación global de capa. A la capa de procesos corresponde la elaboración de reportes que revelan patrones novedosos sobre la situación de las finanzas públicas en El Salvador. A la capa de servicios se asocia la facilitación del acceso y la interacción de los usuarios con los datos de las finanzas públicas. Y finalmente, la capa de infraestructura se monitorea y evalúa en función de la disponibilidad de herramientas tecnológicas innovadoras para la exploración e interacción con los datos de las finanzas públicas.

La plataforma de datos incluirá un módulo adicional de estadísticas con las cuales se elaborarán reportes mensuales de monitoreo. La evaluación de procesos o de resultados se programará en concordancia con el calendario global del proyecto. Se enfatiza que el propósito del monitoreo y la evaluación es el aprendizaje y la mejora continua. Eso incluye, en forma recursiva, la misma metodología de monitoreo y evaluación, la cual se irá mejorando en la medida que la plataforma de datos vaya creciendo.

9 Referencias

Concha, G. y Naser, A. (2012). *Datos abiertos: un nuevo desafío para los gobiernos de la región*. Santiago de Chile: Instituto Latinoamericano y del Caribe de Planificación Económica y Social (ILPES). Serie Gestión Pública No. 74.

Dener, C. & Young, S. (2013). *Financial Management Information Systems and Open Budget Data*. Washington D.C.: The World Bank.

Kitchin, R. (2014). *The Data Revolution*. SAGE.

Moine, J., Haedo, A. y Gordillo, S. (2011). Estudio comparativo de metodologías para minería de datos. La Plata, Argentina: Universidad Nacional de La Plata. *XIII Workshop de Investigadores en Ciencias de la Computación*.

Shafique, U. y Qaiser, H. (2014). A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12(1), 217-222.

Solana, M. (2004). Transparency Portals: Delivering public financial information to Citizens in Latin America. En *Innovative Case Studies on Participatory Instruments*. The World Bank, Civil Society Team.