

Estrategia de Datos



Intendencia
Montevideo



Resumen

La **Estrategia de Datos de Montevideo** tiene el objetivo de generar un mayor impacto en los residentes de la ciudad, a partir de una **mejora en el uso de los datos** que permita incorporar una cultura de datos basada en la **transparencia** y la búsqueda continua de **soluciones**

Su diseño se llevó a cabo a partir de cinco ejes centrales:

- **Gobernanza**
- **Calidad**
- **Capacidad**
- **Transparencia**
- **Uso**

Estos cinco ejes, se apoyan en tres principios fundamentales: el **liderazgo**, el **equipo** y la **ciudadanía**.

GOBERNANZA - *Estructura clara, definición de roles y responsabilidades.*

La creciente necesidad de las organizaciones en mejorar su gestión y su estructura de decisión en base a información en tiempo real, impulsa a trabajar de forma integral y sistemática en la generación y procesamiento de datos.

La creación de un **Comité de Datos** que opera de manera transversal en la seguridad, confiabilidad, integralidad, documentación, administración y auditoría de los datos, garantiza una buena gestión de los procesos necesarios y requeridos para tal fin.

Sus funciones principales son el diseño y ejecución de estrategias que aseguren el relevamiento, documentación y publicación de datos apoyando los procesos internos de trabajo y la rendición de cuentas y transparencia hacia la ciudadanía.

CALIDAD - implementar estándares de consistencia, precisión y seguridad

Actualmente se cuenta con una primera versión del **Inventario Único de Datos de la Intendencia de Montevideo**. Este inventario estará disponible para toda la intendencia con información agrupada por departamento, el cual será responsabilidad de los *referentes de datos* de cada departamento su actualización y mantenimiento.

La mejora de las fuentes y sistemas relevados en el inventario de datos, surgirá la prioridad asignada a los conjuntos de datos específicos, contenidos en dicho inventario. La seguridad y confidencialidad de estos datos en los distintos escenarios de uso (registro, uso interno y uso externo), será provista a partir de la utilización de distintas herramientas y protocolos.

CAPACIDAD - Equipos de trabajo y Plan de Formación y capacitación

En este contexto, es necesario la conformación de equipos de trabajo capacitados y formados para el manejo y análisis de datos.

Para ello, fue diseñado un **Plan de Formación** con objetivos y cronograma detallado, en base al **Relevamiento de Capacidades y Necesidades de los Departamentos** de la intendencia.

TRANSPARENCIA - Disponibilizar los datos interna y externamente

La estandarización y sistematización del **tratamiento y publicación de los datos a partir de protocolos** que toman como referencia a la normativa vigente, permite aportar transparencia al conjunto de procesos que acompañan al ciclo de vida de publicación de los datos.

En este proceso se toman como referencia, el marco legal, los propios datos y metadatos asociados a éstos, la preparación de los datos para su publicación y la publicación de los datos.

Asociado a los procesos de protocolos y tratamiento de los datos, la publicación de datos abiertos requiere de la **pauta para la publicación de datos abiertos**. Esta pauta permite establecer claramente el procedimiento a partir del cual se publica un conjunto de datos. Este procedimiento establece:

1. relevar el requerimiento (brindar espacios y canales de solicitud de datos)
2. procesar el requerimiento (a partir de un registro de incidencias)
3. Definir acción (para satisfacer la necesidad de datos relevada)

Finalmente, la publicación de datos en los observatorios también ofrece la posibilidad de libre consulta de datos. Esta herramienta de análisis y visualización, requiere de las siguientes tareas asociadas a distintos equipos de trabajo:

1. definición de indicadores
2. definición del tipo de visualización requerida
3. prototipación de la visualización a partir de los datos existentes
4. pasaje a producción del producto de los procesos anteriores

USO - Toma de decisiones informadas y mejorar el uso de datos

Los usos asociados a la utilización intensiva de datos para mejorar las políticas públicas y la toma de decisiones, es incalculable. No por el hecho de no poder enumerar las posibilidades actuales, sino más bien, por el hecho de no poder delimitar a mediano y largo plazo, el desarrollo potencial de la utilización de datos en la gestión de organismos públicos.

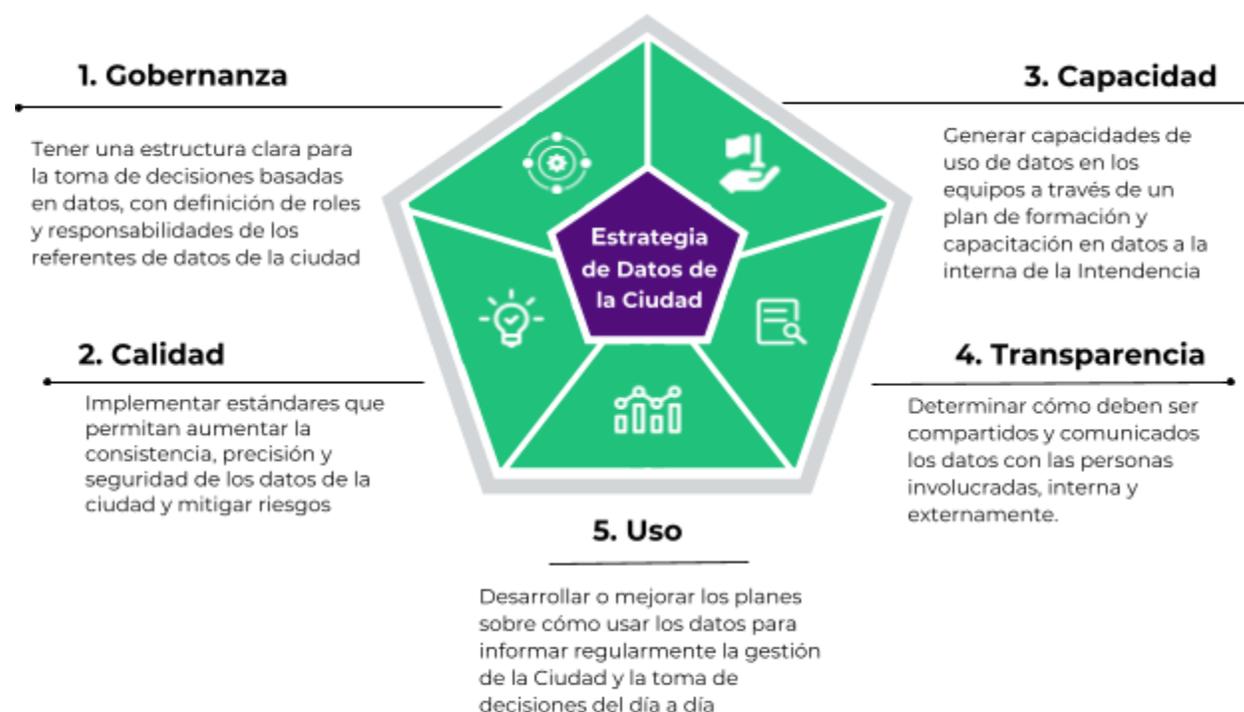
Por el momento, a nivel general, se han propuestos los siguientes usos para la Intendencia de Montevideo:

- Plataforma de integración de datos provenientes de distintas fuentes (sistemas, planillas, sensores, etc.)
- Tablero de mando para prioridades de la intendencia y los departamentos

Introducción

Este documento sintetiza la **Estrategia de Datos de la Intendencia de Montevideo**. En ella se presentan prácticas y protocolos para **mejorar el uso de datos** con el objetivo de generar **mayor impacto en los residentes** de la ciudad.

El diseño de la Estrategia de Datos fue elaborado a partir de cinco componentes.



Para que esta estrategia se traduzca en una mejora tangible de calidad de vida de los residentes de la ciudad, es necesario consolidar una **cultura de datos** sostenible de Montevideo, que permita incrementar la **transparencia** y **rendición de cuentas** de la Intendencia hacia la ciudadanía y co-crear **soluciones** para los problemas públicos.

La cultura de datos de la Intendencia se apoya en tres principios fundamentales: **liderazgo, equipo y ciudadanía**

Liderazgo: *Gestionar con el ejemplo*

- Decisión expresa de Intendenta de usar datos para tomar decisiones diarias como

medio para mejorar los servicios públicos.

- Seguimiento constante a la ejecución de acciones estratégicas definidas por datos.

Equipo: *Retar y reconocer*

- Proponer constantemente nuevos retos.
- Dar estabilidad a los espacios de trabajo.
- Crear espacios de continuo aprendizaje.
- Buscar mejoras y oportunidades continuas para publicar y compartir datos.

Ciudadanía: *Transparencia y comunicación*

- Democratizar el acceso a los datos como datos abiertos.
- Contar con una ciudadanía informada que demanda acceso a datos.
- Incorporar opiniones de la ciudadanía en el diseño e implementación de servicios públicos.
- Generar confianza en los datos hacia la ciudadanía.

A partir de estos cinco componentes principales y los tres principios rectores, se promueven prácticas, rutinas y procesos de trabajo concretos que permiten avanzar en la construcción de herramientas de análisis y uso de datos que ayuden a transformar la realidad de las personas.

Tabla de contenidos:

1. Gobernanza
2. Calidad
3. Capacidad
4. Transparencia
5. Uso



1. Gobernanza

Tener una estructura clara para la toma de decisiones basada en datos, con definición de roles y responsabilidades de los referentes de datos de la ciudad.

1.1 Comité de Datos

“Los datos impulsan la economía de la información”¹.

Hoy en día, el contar en organizaciones como la Intendencia de Montevideo con un grupo de trabajo multidisciplinario, con una visión general de la organización y de toda la información producida en ella, es clave para impulsar una gestión integral.

Actualmente, tener información en tiempo real es el insumo necesario para la toma de decisiones basadas en evidencia y el punto central de una buena gestión. Siempre teniendo a los habitantes de nuestros territorios en el centro y realizando acciones para mejorar la calidad de los servicios y enfrentar los desafíos más importantes.

La importancia del gobierno del dato

El gobierno del dato es un conjunto de procesos que deben ser definidos y gestionados en las empresas u organizaciones para garantizar la seguridad, la confiabilidad, la integridad, la documentación, la administración y la auditoría de los datos. Incluye políticas y procedimientos internos de control de los datos y contribuye a implementar un modelo gestionado de calidad de los datos en todas las fases del ciclo de vida de estos.²

Objetivo

El objetivo de contar con un Comité de Datos pasa por definir y ejecutar estrategias de publicación de los distintos juegos de datos producidos, así como por contar con un catálogo que identifique y documente las características de cada uno de los datos que se manejan en los distintos procesos desarrollados en la Intendencia.

¹ <https://iapp.org/about/#>

² <https://blogvisionarios.com/articulos-data/gobierno-de-datos-importancia-componentes-estrategias-roles/>

El Comité deberá establecer las prácticas y políticas documentadas en torno a la gestión de los datos, tanto internos como externos (aquellos que surgen de los dispositivos colocados en la ciudad) o también los producidos por terceros.

Asimismo, debe ser el responsable de describir las políticas en cuanto al almacenamiento, la definición de estándares y la forma en que se le dará mantenimiento a la información producida por esos datos, y en algún sentido validar la construcción de herramientas para su explotación.

Cometidos

El Comité de Datos deberá:

- Definir, a través de un catálogo o inventario de datos, la lista de datos manejada por la Intendencia y asegurar su mantenimiento.
- Monitorear la implementación y el proceso de cumplimiento relacionado con el uso de datos e información.
- Generar los protocolos para la disponibilización de los datos, tanto en la interna de la Intendencia así como para terceros.
- Asegurar la calidad de los datos (al menos de los que es responsable la Intendencia).
- Establecer roles y responsabilidades en torno a la gestión integral de los datos.
- Proponer revisiones periódicas a las políticas y procedimientos adoptados por la Intendencia en materia de protección de la información, tanto desde el punto de vista físico como de la exposición de datos personales.
- Tener rutinas establecidas para discutir prioridades en materia de datos de la Intendencia.
 - Reuniones periódicas del Comité de Datos (una vez al mes).
 - Envío de Informes de avances y reuniones de presentación a la Intendenta (cada 2 o 3 meses).
- Definir y medir, de manera periódica, el avance y cumplimiento de los objetivos trazados por el Comité a través de indicadores de éxito.
- Revisar y actualizar, de manera periódica, la Estrategia de Datos.

Referentes de datos

Cada Departamento de la Intendencia nombrará a uno o más representantes para su integración al Comité de Datos.

Los referentes cumplirán con las siguientes funciones:

- Ayudarán al cumplimiento de los objetivos del Comité.
- Facilitarán la disponibilidad de los datos propios de sus competencias.
- Establecerán qué datos de otros departamentos son necesarios para realizar sus análisis.



2. Calidad

Implementar estándares que permitan aumentar la consistencia, precisión y seguridad de los datos de la ciudad y mitigar riesgos

2.1 Relevamiento de inventario de datos y sistemas de información

Gracias a los insumos brindados por el relevamiento de sistemas y fuentes de información, organizado por el Departamento Desarrollo Sostenible e Inteligente (DSI) en el marco de la certificación *What Works Cities* (WWC) de la Intendencia de Montevideo, y el relevamiento de datos generados en los siete departamentos seleccionados para la investigación en el marco del programa de *City Data Alliance*, se cuenta con una primera versión del Inventario único de datos de la Intendencia de Montevideo.

El Inventario de datos será de libre acceso para toda la Intendencia. Se pondrá a disposición en un sitio compartido donde se presenta la siguiente información agrupada por departamento:

- Dato que genera: Descripción del dato. En el caso que este dato se publique a la ciudadanía, en la descripción del dato se mantendrá un enlace a la publicación del mismo.
- Origen del dato: Para poder procesar el dato y generar nuevas visualizaciones o indicadores es necesario saber donde se aloja dicho dato. Esta información se registra a través de los siguientes campos en la tabla de inventario:
 - Sistema: En el caso por ejemplo de que la fuente que genera los datos es un Sistema de Software, se debe indicar el nombre del Sistema asociado. Si es otro tipo de fuente se puede indicar alguna referencia de donde se aloja la misma (por ejemplo si es una planilla).
 - Tipo de Fuente: Este campo tiene valores fijos y su objetivo es indicar el tipo de fuente asociada, ejemplo si es una planilla, un Sistema o un sensor. O sea, en este campo se indica que tipo de fuente de datos se tiene y en el campo Sistema se brinda más detalle de la fuente. Los valores posibles son:
 - ◆ Sistema IM: Sistema en la infraestructura de la Intendencia
 - ◆ Análisis de datos: Una de las funciones de la Unidad de Análisis de Datos (la cual depende de Desarrollo Sostenible e Inteligente) consiste en tomar datos de distintas fuentes y procesarlos. En este

procesamiento, se pueden generar nuevos datos. Por ejemplo, a partir de los distintos viajes que realizan las personas en ómnibus y sus subidas se puede inferir donde se bajan las mismas. La fuente que genera esos datos en este caso es Análisis de Datos.

- ◆ Planilla: datos mantenidos en planilla
 - ◆ Dispositivos: datos generados por sensores, cámaras, etc.
 - ◆ Sistema Externo: datos generados por Sistemas externos (incluye Sistemas donde no somos dueños de los datos como es el caso de ASSE, o Sistemas cuya infraestructura no lo maneja la Intendencia)
 - ◆ Base de datos: datos que no tienen Sistema asociado pero si hay una base de datos, ejemplo base de datos de Sucive
- Responsable del dato: responsable o dueño del dato, o sea, es el área o la persona referente del dato (para el caso que se necesite realizar una consulta).
 - Frecuencia de generación: frecuencia en la que se generan los datos
 - Información adicional. Otra información adicional que se necesite documentar

Es responsabilidad de los referentes de datos de los departamentos mantener actualizado el inventario.

Los registros serán documentados con el mayor nivel de granularidad posible, es decir, se mantendrá una entrada en el inventario para cada uno de los distintos datos que gestiona o utiliza cada departamento.

A continuación se listan los posibles escenarios en los cuales se deberá actualizar el inventario de datos de un Departamento.

- Se generan nuevos datos dentro del Departamento, ya sea por la implementación de un nuevo programa, la puesta en producción de un nuevo Sistema de Software etc.
 - Se debe generar una nueva entrada en el inventario y completar los datos de la tabla
- Cambia la fuente asociada de un dato, la frecuencia de actualización o el responsable del dato.
 - Se actualiza el dato correspondiente en la tabla
- Se publica a la ciudadanía un dato del inventario
 - En el caso que un registro en el inventario se publique a la ciudadanía, desde el nombre de dato se debe mantener un enlace a la información publicada.

2.2 Mejora de la calidad de los datos de alto valor

Con el objetivo de mejorar la gestión de los datos se identificarán los sistemas y fuentes más vulnerables siguiendo los criterios definidos. Del cruce de los sistemas vulnerables con los datos prioritarios surgirá una propuesta de cronograma de mejora de los sistemas y fuentes de información.

Estas actividades serán coordinadas y realizadas por los referentes del Comité de datos, y articuladas con los demás actores clave relacionados a los procesos de mejora de sistemas en la institución.

2.3 Seguridad en el uso de Datos

En el caso de uso de datos uno de los principales factores a considerar es el tema de su seguridad, principalmente en relación a lo vinculado como datos sensibles o personales.

Los escenarios a considerar son:

- Registro o ingreso de datos
- Uso de los datos dentro de la organización
- Uso de los datos externo a la organización

Registro o Ingreso de datos

Se deberá dar a conocer al ciudadano cuando se recolectan sus datos, claramente el propósito de su uso así como asegurar la confidencialidad de su información.

Instancias donde se identifica el registro de los datos:

- Formularios web
- Aplicaciones móviles
- Entrevistas
- Formularios en papel, entre otros.

Uso de datos dentro de la organización

Los datos recolectados por cualquier vía, podrán ser utilizados dentro de la organización siempre y cuando se asegure la confidencialidad de datos personales incluidos en las excepciones de acuerdo a la Ley N° 18.381 de Acceso a la Información Pública.

Si correspondiere, es necesario definir un procedimiento que permita compartir información entre los departamentos y dependencias pero que asegure que tales datos serán preservados como indica la ley mencionada

Uso de datos externo a la organización

La Intendencia no podrá compartir sus datos libremente. No obstante, la forma de compartirlos puede ser de 2 maneras:

- En formato de dato abierto
- A través de acuerdo con otras instituciones

En cualquier caso, no puede compartir datos que estén regulados por la Ley N°18.331 anteriormente mencionada.

En el caso de publicación de datos en formato abierto (csv, txt, otro) éstos quedan disponibles en el Catálogo Nacional de Datos Abiertos (<https://catalogodatos.gub.uy/>) y próximamente en nuestro propio sitio institucional (<https://ckan.montevideo.gub.uy/>). En estos casos, los datos están licenciados bajo la [Licencia de Datos Abiertos de Uruguay](#).

En el caso de compartirlos a través de algún acuerdo de cooperación, se deberá firmar un acuerdo de confidencialidad entre las partes en cuanto a la preservación de los datos amparados por la ley.



3. Capacidad

Generar capacidades de uso de datos en los equipos a través de un plan de formación y capacitación de datos en la Intendencia.

Para la elaboración del plan de formación y capacitación de datos, se tomó como insumo el análisis realizado en base al cruce de necesidades vinculadas a datos y las capacidades detectadas en las distintas áreas. Estos dos insumos son fruto de un trabajo de relevamiento del Comité de Datos, realizado en siete departamentos de la Intendencia.

3.1 Relevamiento de necesidades y capacidades de departamentos

A continuación se presentan las conclusiones principales de dicho análisis:

- En base al relevamiento de capacidades realizado en 7 departamentos, se concluye que hay personas que dicen estar familiarizadas o que utilizan por lo menos una herramienta para el análisis de datos.
- En todos los departamentos relevados, al menos un 18% de las personas declara tener conocimientos de lenguajes de programación.
- Si bien la mayor parte de las personas consultadas afirma extraer datos de bases de datos institucionales para su trabajo habitual, generalmente los mismos no se procesan de forma sistemática (conclusión que se desprende por los tipos de herramientas que se usan y del escaso uso de herramientas de tablero básicas).
- La percepción de necesidades vinculadas a datos en los departamentos arroja que las mismas se pueden agrupar en 5 grandes categorías:
 - ◆ Capacidades y habilidades
 - ◆ Personas dedicadas al análisis de datos
 - ◆ Protocolos
 - ◆ Plataformas y sistemas
 - ◆ Cultura y prácticas

Siendo la falta de protocolos y cultura de datos las preponderantes.

- Según el análisis realizado sobre los resultados obtenidos, la percepción de necesidades en capacidades y/o en personas dedicadas al análisis de datos no se correlaciona con sus recursos existentes.
- Para trabajar sobre estas necesidades y en base al relevamiento de capacidades se proponen como líneas de trabajo generales:

- ◆ Continuar con las capacitaciones en herramientas para el uso y análisis de datos, focalizando en los perfiles sobre los que cada departamento quiera potenciar.
- ◆ Creación de la figura de referentes de datos en cada departamento que promueva la cultura de datos e institucionalización de la misma.
- ◆ Definición de protocolos cuando no existan o institucionalización, en algunos casos, de los existentes (ver punto 4.1).

Cuadro NI.
Comparativo entre la percepción de necesidades vinculadas a capacidades y análisis de datos contra capacidades relevadas

Depto	Percepción de necesidades		Capacidades relevadas de manejo de herramientas (cant. de personas)				
	Capacidades y habilidades	Personas dedicadas al análisis de datos	Bases de datos	Análisis de datos	Business Intelligence	Georreferenciación	Lenguajes de prog.
DA			12	16	12	47	22
Mov		X	5	4	5	22	16
Planif	X		7	11	4	47	11
GH	X		8	4	5	4	14
DS			8	17	13	1	23
Cul	X	X	0	3	7	4	9
DU		X	8	6	6	28	14
Total	3	3	48	61	52	153	109

3.2 Referentes de datos

Se identificarán las habilidades de los referentes de datos y el personal en general en torno al tratamiento y publicación de datos, con el objetivo de considerarlos como potenciales creadores y docentes de cursos específicos al resto del personal de la Intendencia que se enmarquen en el Plan de capacitaciones en datos.

Asimismo se explorarán los convenios vigentes de la organización con universidades para brindar aquellas capacitaciones en las que la IM no encuentre en su personal el conocimiento necesario para llevarlas a cabo.

3.3 Plan de desarrollo de habilidades para la gestión de datos

En base al relevamiento de necesidades y capacidades de departamentos, se desarrolló el Plan de desarrollo de habilidades para la gestión de datos 2023.

Cuadro N2.
Detalle del Plan de desarrollo de habilidades para la gestión de datos

NIVEL DE LA ESTRATEGIA	OBJETIVO / ASPIRACIÓN	POBLACIÓN OBJETIVO	ENTREGABLE 2023	IMPACTO ESTIMADO
NIVEL 0 – INDUCCIÓN Estrategia de datos de la IM Protección de Datos Personales Acceso a Información Pública Seguridad Informática	Desarrollar un programa de inducción al trabajo con datos, con cobertura progresiva, entre el personal que trabaja con computadora.	Personal que trabaja con una computadora.	Diseño del Curso de Inducción en Datos que incluya el Módulo de política de datos institucionales (MPDI). 20 cursos ofrecidos.	500 personas
NIVEL 1 – ALFABETIZACIÓN Nivel I ¿Qué quiero saber? ¿Qué datos necesito obtener? Protocolización del registro, documentación y obtención de datos	Lograr que al menos el 50% del funcionariado priorizado por cada Departamento acceda al curso.	Personal priorizado por cada departamento, entre quienes registran datos.	Diseño del Curso de Registro y Obtención de Datos Institucionales (RODI). 20 cursos ofrecidos.	500 personas
NIVEL 1 – ALFABETIZACIÓN Nivel II Conformación y depuración de bases de datos Análisis Primario	Lograr que al menos el 25% del funcionariado priorizado por cada Departamento tenga conocimientos en 2 herramientas seleccionadas.	Personal priorizado por cada departamento, entre quienes analizan datos.	Oferta de 10 cursos entre las distintas herramientas de análisis de datos priorizadas.	250 personas
NIVEL 2 – PROYECTOS DE DATOS Análisis intermedio/avanzado Preparación de datos para el Observatorio Visualizadores y productos comunicacionales	Lograr que al menos el 25% del funcionariado priorizado por cada Departamento tenga conocimientos en 1 herramienta seleccionada.	Personal priorizado por cada departamento, entre quienes realizan análisis intermedio de datos.	Oferta de 4 cursos entre las distintas herramientas de análisis de datos priorizadas.	100 personas
NIVEL 3 – ANÁLISIS DE DATOS AVANZADO Machine Learning	Lograr que al menos el 25% del funcionariado priorizado por cada Departamento tenga conocimientos en 1 herramienta seleccionada.	Personal priorizado por cada departamento, entre quienes realizan análisis avanzado de datos.	Oferta de 1 curso entre las distintas herramientas de análisis de datos priorizadas.	20 personas



4. Transparencia

Determinar cómo deben ser compartidos y comunicados los datos con las personas involucradas, interna y externamente.

4.1 Protocolo de proceso de tratamiento y publicación de datos

Objetivo

Esta sección pretende introducir al tema del ciclo de vida de publicación de los datos, y tiene como objetivo estandarizar y sistematizar el proceso de tratamiento y publicación de datos en la organización teniendo en consideración la normativa vigente.

Los beneficios que se espera lograr con la aplicación de este protocolo, son los siguientes:

- Comprensión de la estructura y significado de datos publicados
- Facilidad de procesamiento de los datos
- Facilidad de descubrimiento de datos por parte de terceros interesados
- Reutilización de los datos
- Confianza en los datos por parte de los usuarios
- Interoperabilidad

Marco Legal

La ley 18.381 sobre el derecho de acceso a la información pública, “tiene por objeto promover la transparencia de la función administrativa de todo organismo público, sea o no estatal, y garantizar el derecho fundamental de las personas al acceso a la información pública”, definiendo esta como aquella información producida, obtenida, en poder o bajo control de los organismos públicos (artículos 1 y 2).

La información mínima que los organismos públicos deben difundir, según el artículo 5, es:

- A) Su estructura orgánica.
- B) Las facultades de cada unidad administrativa.
- C) La estructura de remuneraciones por categoría escalafonaria, funciones de los cargos y sistema de compensación.
- D) Información sobre presupuesto asignado, su ejecución, con los resultados de las auditorías que en cada caso corresponda.
- E) Concesiones, licitaciones, permisos o autorizaciones otorgadas, especificando los titulares o beneficiarios de éstos.

- F) Toda información estadística de interés general, de acuerdo con los fines de cada organismo.
- G) Mecanismos de participación ciudadana, en especial domicilio y unidad a la que deben dirigirse las solicitudes para obtener información.

Es importante considerar que las excepciones a la información pública comprenden “aquellas definidas como secretas por la ley y las que se definan seguidamente como de carácter reservado y confidencial” (artículo 8). Dentro de la información confidencial se encuentran los datos personales, para lo cual hay que remitirse también a lo establecido en la ley 18.331 de protección de datos personales.

Ciclo de Vida de los Datos (datos y metadatos)

La organización debe contar con infraestructura / herramientas y un proceso adecuado para la gestión de la publicación de datos.

Un componente importante de este proceso es el portal de publicación de datos, que consiste en un aplicación web dedicada específicamente a la organización, búsqueda y descarga de los conjuntos de datos.

El otro concepto relevante es el del rol de dueño o responsable de los datos, que tiene distintas acciones dentro del proceso / protocolo que se describe en este documento.

El protocolo tiene las siguientes etapas:

- Ingresar los datos en la plataforma desde los distintos servicios internos
- Preparar la información para su publicación
- Publicación de los datos en el portal de publicación

Ingreso de los datos

Según la naturaleza del dato será distinto el tratamiento para ingresarlo en la plataforma de gestión. Se consideran 2 grandes grupos principales:

- Archivos de texto (CSV, XLS, TXT, XML, JSON, ... etc.)
- Repositorios de datos

Los datos en archivos tendrán su ingreso al sistema a partir de una carpeta compartida específica en el sistema de archivos de la organización, en la cual el dueño de la información dejará el archivo con un determinado formato para el nombre.

En el caso de repositorios de datos existentes el acceso se realiza mediante mecanismos de consulta de estos repositorios, como puede ser a través de SQL, servicios REST, etc. Estos accesos deberán utilizar roles / permisos de solo lectura, es decir, que no se pueda modificar datos o estructuras de datos en el repositorio origen.

Registro de metadatos

El rol del dueño de los datos es de quien se toman las siguientes definiciones de metadatos.

Se cuenta con 2 tipos de metadatos:

- Definiciones operativas
- Definiciones funcionales y/o de semántica de los datos

Las definiciones operativas se deben realizar como parte de la etapa de ingreso y la definición funcional puede hacerse en cualquier momento del proceso, pero se considera una condición necesaria para la publicación, es decir, no se publicará en forma definitiva el conjunto de datos hasta no contar con estos metadatos.

Las definiciones operativas consisten de información relativa al correcto procesamiento de los datos, donde se incluye:

- Ventana de tiempo de los datos
- Periodicidad de la actualización
- Dónde queda almacenado para la publicación
- Referentes para comunicación de eventos relativos a la publicación

Algunas de ellas son específicamente internas a la plataforma de procesamiento y por lo tanto las decide el personal técnico, las restantes se definen consultando a los dueños de los conjuntos de datos.

Las definiciones funcionales describen a los datos, quien los mantiene, etc. Se deberán ingresar por parte de los dueños de los datos a través de accesos con roles / permisos adecuados a la sección organizacional correspondiente en el portal de publicación.

El portal está estructurado en categorías que agrupan / organizan los conjuntos de datos y en el concepto de publicadores, que se corresponde con el de responsables de una o varias categorías. El publicador tiene los roles para poder modificar los metadatos de los conjuntos de datos de sus categorías.

Preparación de los datos

Esta etapa consiste en mover efectivamente los datos desde su interfaz de ingreso correspondiente (de acuerdo a su naturaleza, como se especificó en la etapa de ingreso) a un destino interno en la plataforma adecuada para su publicación.

Aquí opcionalmente se pueden aplicar pasos de validación y/o formateo de los datos para garantizar su correcto procesamiento o estado desde el punto de vista sintáctico.

El destino de los datos pueden ser carpetas o repositorios específicos internos de la plataforma, dependiendo en cada caso por aspectos técnicos como rendimiento, modo de acceso a los datos publicados, etc.

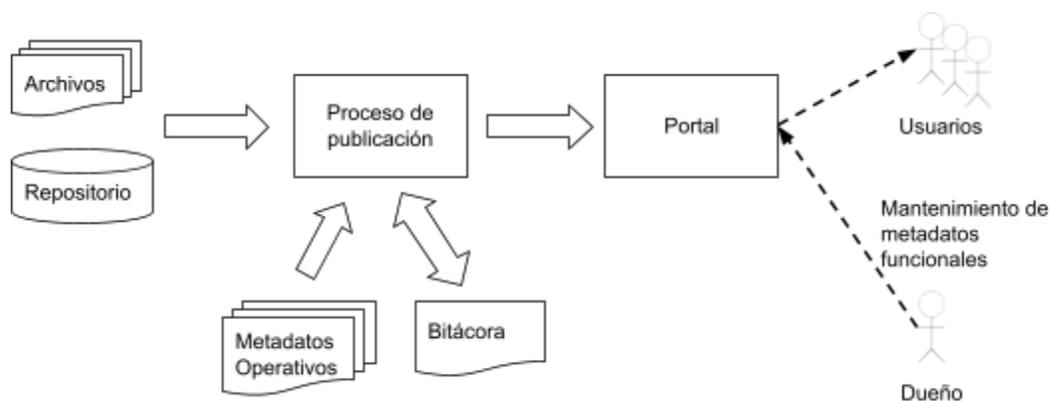
Todo el procesamiento queda registrado en una bitácora, que permite tener la trazabilidad de las publicaciones, detección de procesamiento repetido / retoma, notificación final de la publicación / no publicación debido a errores, etc. Esa bitácora incluirá la fecha del procesamiento, el origen y el resultado de la publicación.

Publicación de los datos

En esta última etapa se aplica un proceso automatizado para subir los datos al portal.

En este momento cada conjunto de datos definido fue ingresado y procesado en la plataforma por lo cual existirá un archivo o consulta a determinado repositorio en donde reside para ser publicado. De acuerdo a la frecuencia de actualización definida en los metadatos, existe un procesamiento batch automático que utiliza el API del portal para realizar la actualización sin interacción humana.

Diagrama N1.
Procesamiento batch automático



Al finalizar la publicación se enviará una comunicación a los responsables del conjunto de datos y a los grupos técnicos con la notificación del resultado, es decir, si quedó efectivamente publicado o si hubo un fallo y en ese caso, de ser posible, la causa del mismo. Esta información se obtiene / actualiza de la bitácora de procesamiento de datos.

En términos generales, existen 2 grandes familias de fallos:

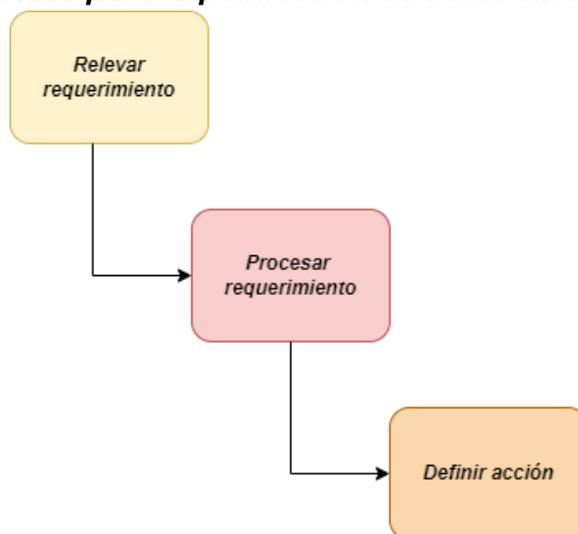
- Fallo funcional (no pasar la validación sintáctica / semántica)
- Fallo técnico (indisponibilidad de algún recurso en la plataforma, etc.)

En el caso del fallo técnico, se puede reintentar la publicación cuando se resuelvan los problemas sin necesidad de solicitar acción por parte del dueño de los datos.

En el fallo funcional, el dueño de los datos deberá enviar un conjunto corregido o al menos comunicarse con los referentes técnicos para solucionar el tema.

4.2 Pauta para la publicación de datos abiertos

Diagrama N2.
Proceso para la publicación de datos abiertos



Relevar requerimientos sobre la publicación de datos.

El primer paso en la publicación de datos abiertos es poder brindar, tanto a la ciudadanía como al personal de gobierno, distintos espacios donde puedan plantear sus dudas, comentarios o nuevos requerimientos.

Para el caso de la comunidad, la Intendencia ofrece varios canales con el fin de solicitar comentarios con respecto a los datos:

- Portal de datos abiertos:
 - En este portal existe un formulario donde las personas pueden hacer comentarios o sugerencias.
 - Para cada conjunto de datos, se indica el contacto del mail del responsable de ese conjunto. Las personas pueden contactar al responsable y plantear sus dudas o sugerencias.
- Buzón ciudadano - https://montevideo.gub.uy/formularios/buzon_ciudadano
- Formulario de acceso a la información.
 - A partir de dicho formulario el ciudadano o ciudadana puede solicitar los datos que necesite. A modo de ejemplo, se puede solicitar información que tenga la Intendencia de su persona y solicitar la corrección correspondiente en caso de ser necesario.
<https://formularios.montevideo.gub.uy/solicitud-acceso-informacion-publica>
- Formulario desde el sitio de la Intendencia
<https://montevideo.gub.uy/formularios/solicitud-de-nuevos-datos-o-servicios>

A partir de los canales mencionados, es que la Intendencia recibe las distintas solicitudes, consultas o inquietudes por parte de la ciudadanía relacionados a distintos temas en los cuales se incluye problema de datos. Las mismas son canalizadas al área o responsable correspondiente los cuales son notificados vía mail.

Si el requerimiento surge del personal de Gobierno, se dispone de la casilla *datosabiertos@imm.gub.uy*. Esta casilla es atendida por todo el equipo asociado a la publicación de datos abiertos.

Procesar el requerimiento

En todos los casos, el equipo de Datos Abiertos coordina reuniones con la persona o área responsable de los datos con el objetivo de analizar cómo se generan los datos, la fuentes asociadas, si es posible publicar el datos de forma sistemática, periodicidad asociada etc.

Para llevar registro de este trabajo se mantendrá un registro en el Sistema de Incidencias.

Definir la acción a tomar

A partir del análisis mencionado pueden surgir las siguientes acciones:

1. Agregar un nuevo conjunto de datos
2. Se vuelva a publicar un set de datos
3. Cambiar la metadata del conjunto de datos o recursos si fuera necesario
4. No realizar ninguna acción

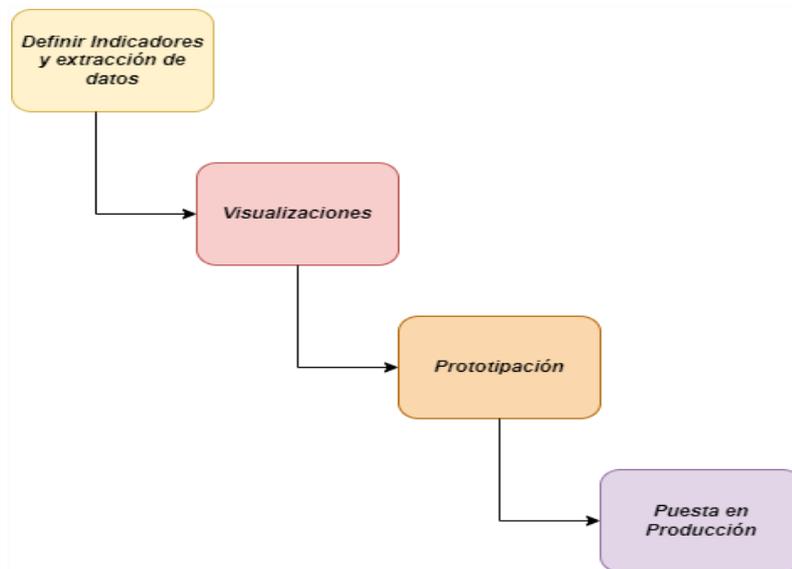
En todos los casos se debe dar respuesta a la persona de las acciones tomadas.

Para la publicación de datos abiertos se cuenta con los documentos de pautas.

4.3 Pauta para la publicación de datos en los observatorios

El desarrollo de un nuevo observatorio temático se realiza en conjunto entre los responsables funcionales del Departamento correspondiente y la Unidad de Análisis de Datos y Plataforma e integrantes del Comité de Datos.

Diagrama N3.
Proceso para la publicación de datos en los observatorios



El primer paso es definir los **indicadores** que se desean publicar y por cada indicador se debe determinar la fuente asociada de datos, ya que dependiendo de esta es como se procederá con la **extracción de datos**.

Por más detalles, ver de los tipos de fuentes ver documento [Montevideo] Estrategia de datos para toda la ciudad/5.1 Plataforma de integración de datos

Teniendo definido cómo será el procesamiento de datos se deben definir las **visualizaciones** que se quieren tener. Se pueden ver ejemplos de visualizaciones en el Observatorio Ambiental <https://ambiental.montevideo.gub.uy/>

Teniendo definido como es la extracción de datos y cómo éstos se visualizarán, empieza una etapa de **prototipación** donde se detectan problemas de datos, se habla con el área responsable y se ajusta el proceso.

Una vez que la etapa de prototipación culmine y el departamento responsable dé su aprobación de las distintas visualizaciones del observatorio, se procede con el **pasaje a producción**.



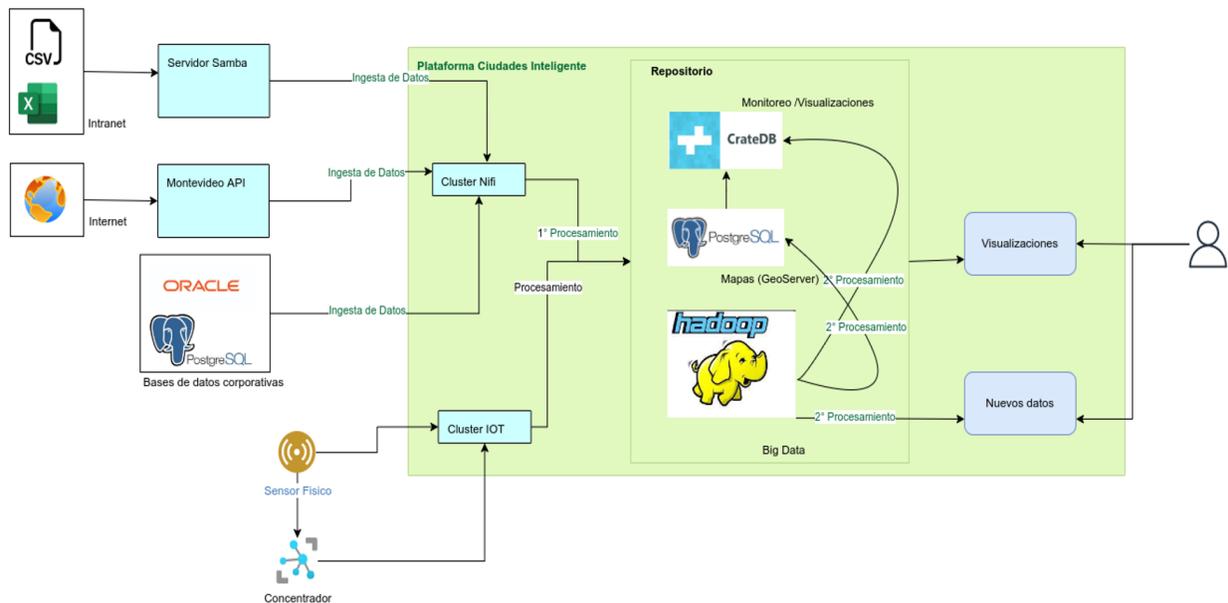
5. Uso

Desarrollar o mejorar los planes sobre cómo usar los datos para informar regularmente la gestión de la Ciudad y la toma de decisiones del día a día.

5.1 Plataforma de integración de datos: “Montevidata”

A continuación se presenta un diagrama cuyo objetivo es mostrar la arquitectura del actual protocolo para la integración de datos:

Diagrama N4.
Arquitectura del actual protocolo para la integración de datos



Es importante aclarar que si bien en el diagrama se detallan los Manejadores de Base de datos como los tipos de archivos que hoy se utilizan, la arquitectura de Integración de Datos puede extenderse a nuevos tipos de archivo como también a otros Manejadores de Bases de Datos.

Fuente de datos

Como se puede apreciar en el diagrama, los datos pueden provenir de distintas fuentes.

Planillas

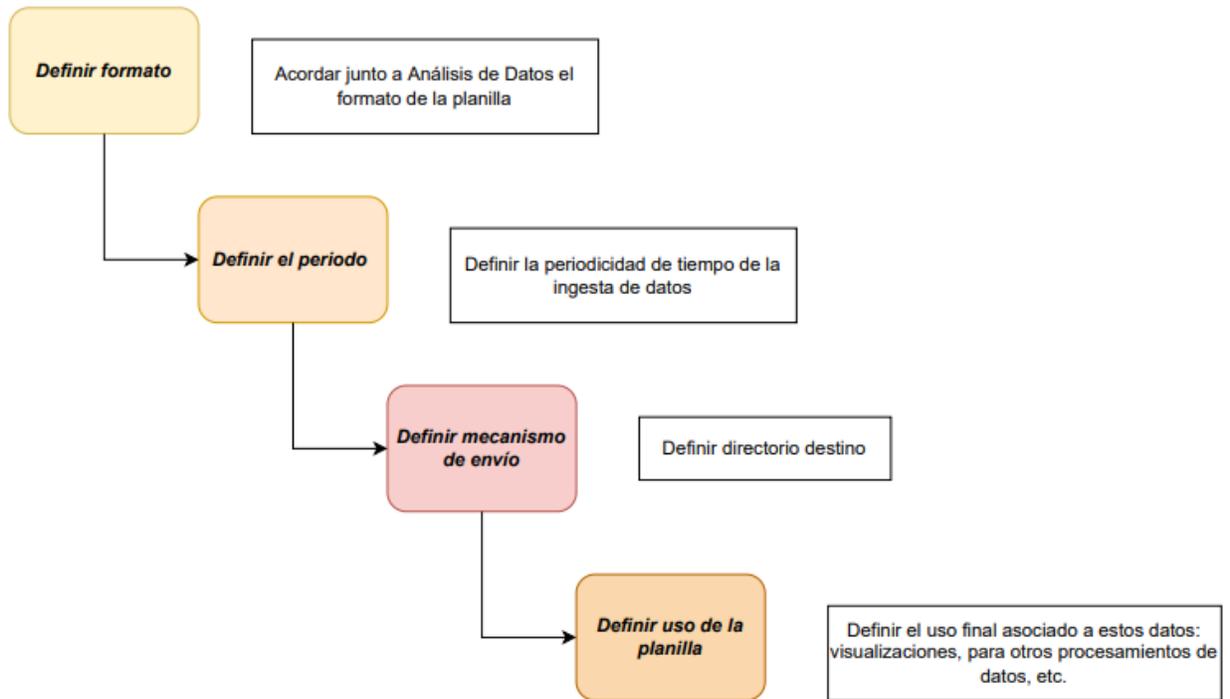
Este escenario ocurre cuando:

- No se tiene un Sistema de Gestión para el registro de datos y como herramienta para el registro se utiliza una planilla.
- Se registran los datos en un Sistema, pero no es posible extraerlos directamente de la base de datos o mediante servicios.

Actualmente el formato que se soporta es excel, calc o csv, pero se puede adaptar la solución para soportar otros formatos.

Diagrama N5.

Consideraciones para incorporar una nueva fuente de datos de tipo planilla



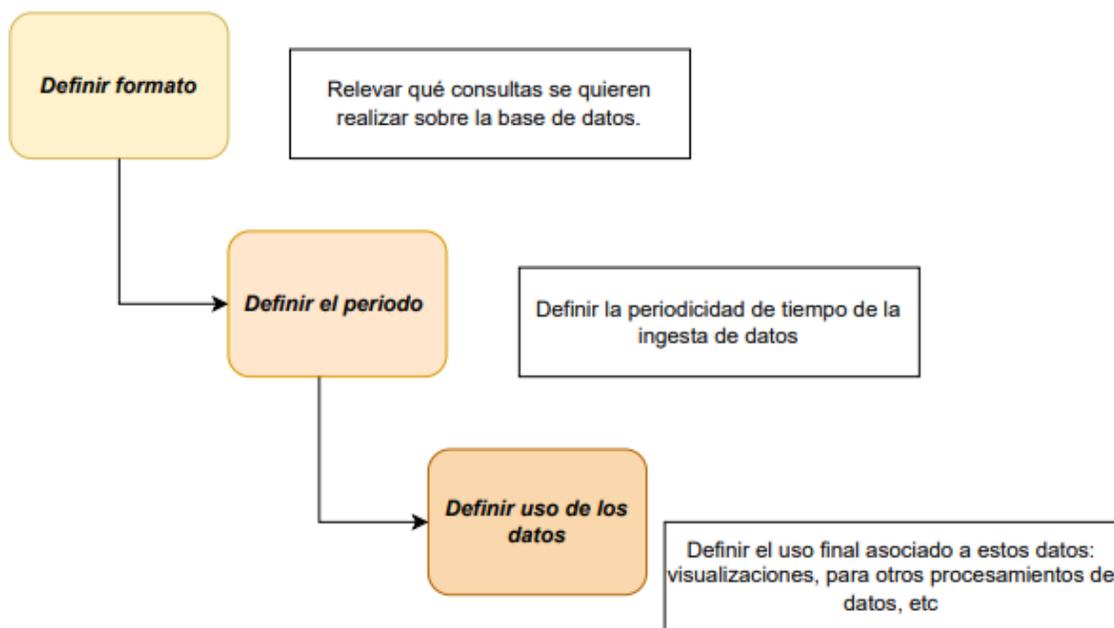
- Acordar su formato, es decir, el nombre de la planilla así como el nombre, orden y tipo de las columnas.
- Definir el periodo en el cual la planilla se encontrará disponible para ser procesada.
- Definir el directorio destino: para que pueda ser procesada la planilla deberá ser copiada a un servidor Samba, a un directorio en particular. En el caso de que el envío del archivo sea a través de internet, la comunicación será utilizando la plataforma MontevideoAPI

Base de datos

Actualmente se están procesando Bases de Datos Oracle, Postgres y SqlServer, las cuales corresponden a bases de datos de Sistemas de Software de la Intendencia de Montevideo. Es importante aclarar que la arquitectura actual permite incluir otros manejadores de Base de Datos con muy bajo costo.

Diagrama N6.

Consideraciones para incorporar una nueva fuente de datos de tipo Base de datos



Para definir las consultas asociadas a la base de datos se deberán mantener reuniones con los responsables del modelo de la Base de Datos, con el fin de entender el mismo y llegar a las consultas correctas.

Sensores

Los datos de sensores en general tienen las siguientes características:

- Son datos generados en un instante de tiempo
- Se refieren a un sensor específico
- Son variables de acuerdo al tipo de sensor

A su vez, según el flujo de datos generado por los sensores se pueden clasificar en 2 grandes grupos: los de gran cantidad de información, ya sea en ráfaga o constante pero en períodos de tiempo muy breves; y otros que tienen un flujo de datos dado en períodos de tiempo mucho más amplios, que podríamos llamar de flujo “estándar”. Para dar ejemplos, el primer caso es el de dispositivos de GPS en vehículos donde se requiere

construir por ejemplo el recorrido. En el segundo caso es el de sensores de temperatura, humedad, etc. Por otra parte el sensor puede comunicarse directamente con el agente o a través de un componente que actúa como concentrador en donde puede haber capacidad de almacenamiento temporal, servicios de consulta, etc.

El protocolo utilizado y recomendado para flujos “estándar” es el protocolo MQTT con formato de datos UltraLight (UL) o en su defecto JSON. El UL es un formato de texto en donde se define cada campo con una sola letra como identificador y el valor que reporta. Para el caso de flujos importantes el listener usa protocolo Rest / JSON.

El equipo de Plataforma de CI proveerá en estos casos la documentación y soporte para utilizar ambos tipos en el caso de agregar nuevos sensores como fuente de datos.

Servicios

En este caso se utilizará la plataforma Montevideo API Ejemplo de uso son: servicios vinculados al transporte público de Montevideo y los que describen el estado de bañabilidad de las playas.

Ingesta y procesamiento de Datos

La ingesta y el procesamiento de los datos de las distintas fuentes (a excepción de los sensores) se realiza dentro de la herramienta de gestión de flujos NiFi de Apache, que básicamente permite coordinar / sincronizar distintos procesos genéricos de datos para generar flujos complejos de procesamiento de información.

Para el caso de los sensores, el responsable de la ingesta y el procesamiento es el componente Cluster IOT.

El flujo que siguen los datos dependerá del tipo de datos y uso que se le quiere dar a éstos.

Por ejemplo, como se puede observar en el diagrama, los datos que son de tipo geográficos se vuelcan primero en una base de Postgres y después de procesados se llevan a la base Crate (a partir de la cual se realizan la mayor parte de las visualizaciones).

Hadoop será responsable de procesar grandes volúmenes de datos y el resultado de este procesamiento puede volcarse a la base Crate, o a Postgres (en el caso de ser datos geográficos). También puede suceder que a partir de este procesamiento se genere un nuevo conjunto de datos que puede ser utilizado por otros departamentos para realizar informes o análisis (por ejemplo utilizando como herramienta R).

Grafana es una de las principales herramientas para realizar distintas visualizaciones, la cual tomará los datos desde Crate.

Otra herramienta de visualización utilizada es Qlik, en este caso los datos que se generan en Hadoop se vierten a Hive y de ahí son tomados por esta herramienta.

Calidad de datos acorde a las fuentes

Dependiendo de la calidad de datos de cada fuente es el esfuerzo que se debe realizar para obtener un buen resultado tanto en las visualizaciones como en la generación de nuevos juegos de datos.

En el marco de este análisis, las dimensiones de calidad que se consideran son las siguientes:

- Exactitud: indica qué tan precisos, válidos y libres de errores están los datos
- Completitud: indica si se tiene toda la información de interés.
- Frescura: indica qué tan actualizados o desactualizados están los datos.
- Consistencia: indica si se cumplen las reglas semánticas entre los datos.
- Unicidad: indica el nivel de duplicación entre los datos.

Cuando existe intervención humana en el registro o manipulación de los datos, la probabilidad de que esta acción afecte de forma negativa alguna de las dimensiones de calidad de estos datos es alta.

Cuando la fuente de datos es una planilla, si el registro es manual las dimensiones de exactitud, consistencia, completitud y unicidad pueden verse afectadas.

En el marco de la solución mencionada anteriormente para la integración de datos, para que la planilla sea procesada debe ser copiada a un servidor. Si este paso también es realizado de forma manual, es altamente probable que existan demoras y que los datos no estén disponibles cuando se necesiten, afectando la dimensión de calidad de frescura.

Las acciones que pueden servir para prevenir los problemas planteados son:

- Agregar controles en la planilla para validar ciertos datos.
- En caso de ser factible evaluar la implementación de un sistema que sustituya el registro por planilla.
- Si la planilla se genera a partir de otro Sistema evaluar sustituir dicho mecanismo de integración de datos por uno que sea vía servicios.
- Definir un proceso automático para la copia de archivos.

Cuando los datos son registrados a través de un Sistema, es deseable que el propio Sistema contenga ciertos controles que eviten que se produzcan problemas de calidad en las dimensiones de exactitud, unicidad y completitud.

Si al procesar los datos se detectan errores en algunas de estas dimensiones, se debe consultar con el responsable del Sistema a fin de realizar la corrección correspondiente. El registro de este problema debe ser realizado en el Sistema de Incidencias de la Mesa de Servicios.

En el caso de procesar datos históricos, es preciso que el Sistema mantenga esa información de forma correcta. Por ejemplo, si existe una relación entre dos entidades como ser proveedor y producto, y dicha relación es importante al momento de registrar una factura; si en cierto periodo ese proveedor deja de brindar ese producto es

importante igual mantener el registro de dicha información en el caso de querer procesar información histórica en relación a la facturación.

Estos casos se detectan al momento del procesamiento de los datos con Hadoop y en el caso de encontrar este tipo de problemas, se deberá ingresar una incidencia en el Sistema de Incidencias de la Mesa de Servicios para corregirla.

En el caso de sensores (o cámaras que no habían sido mencionadas anteriormente), es más probable que ciertas dimensiones se mantengan de forma correcta, como ser frescura, exactitud, consistencia y unicidad. En el caso de falla de algún sensor o pérdida de conectividad, si no se establecen alarmas puede ser que se tenga problemas de completitud ya que se puede perder información en cierto periodo de tiempo. Si esa información se encuentra almacenada de forma local en el sensor, es posible reprocesar los datos.

Integración de datos de distintas fuentes

En el caso de integrar datos de distintas fuentes es importante considerar si las entidades que son comunes (por ejemplo si en los dos sistemas se maneja el campo cédula de una persona) tienen mantienen la misma sintaxis. Para estos casos es importante definir quién es el “dueño” de cada dato y así definir qué criterio seguir.

5.2 Plan de diseño de Tableros de Mando para prioridades

Los tableros de mando constituyen una importante herramienta de comunicación y toma de decisión basada en datos sobre las prioridades de la intendencia y los departamentos. Su principal función es la presentación de información sintética sobre el avance de ejecución de las prioridades máximas, por parte de los actores responsables de su ejecución. Su ámbito de aplicación más frecuente es el Gabinete de Gobierno Departamental y el Gabinete de Departamento.

El tablero de mando está integrado por tres componentes fundamentales que apoyan su función:

- monitoreo de indicadores
- alertas sobre problemas de cumplimiento a tiempo de las metas establecidas y
- balance general de ejecución de los mismos (lineal o ponderando el peso relativo de cada uno).

El componente de monitoreo de los indicadores deberá contar, como mínimo, con la siguiente información:

- Semáforo de nivel de cumplimiento en tiempo del indicador,
- Línea de base con el valor del indicador al momento del inicio del período de gobierno,
- El avance actual del valor del indicador,
- La meta establecida para el presente año y

- La meta establecida para el final de período (o final de la medición).

El componente de toma de decisión contará con alertas sobre problemas de cumplimiento en tiempo de cada uno de los indicadores presentados, estas alertas establecen condiciones propicias para la toma de decisión oportuna e informada por parte de las autoridades responsables del cumplimiento, para asegurar el logro de la meta establecida.

El tercer componente, brinda información sobre el balance general de cumplimiento en tiempo sobre los indicadores presentados. Dicho balance podrá realizarse de forma lineal (asignando el mismo peso relativo a cada indicador) o ponderado (asignando un peso relativo distinto a cada indicador, de acuerdo a un criterio de relevancia). Este balance muestra de forma sintética, el estado de ejecución de las prioridades máximas de la intendencia, de los departamentos y/o Áreas Temática.

Actualmente, se trabaja en el desarrollo de la versión 1.0 de los tableros de mando para su implantación dentro del plan piloto en los Departamentos de Desarrollo Ambiental, Movilidad y Secretaría General con el apoyo del Departamento Desarrollo Sostenible e Inteligente previsto para que entre en funcionamiento en el segundo cuatrimestre del año 2023. El proceso inicial responde a la identificación de los indicadores prioritarios del departamento bajo una caracterización de áreas temáticas y con metodología BID, seguido de la evaluación de las fuentes de datos que proporcionarán la información (dentro del inventario de datos), el desarrollo del entorno *web* de visualización y la disponibilización de los datos requeridos.