



Informe Final Proyecto

Publicación de información relevante de Accidentabilidad de Trayecto de la ACHS bajo estándares internacionales de Open Data (217-2018)

Cristián Rodríguez Arenas

2020-03-20

Este trabajo fue seleccionado en la Convocatoria de Proyectos de Investigación e Innovación en Prevención de Accidentes y Enfermedades 2018 de la Superintendencia de Seguridad Social (Chile) y fue financiado por la Asociación Chilena de Seguridad, a través de la Fundación Científica y Tecnológica (FUCYT-ACHS), con recursos del Seguro Social de la Ley N°16.744 de Accidentes del Trabajo y Enfermedades Profesionales.

Antecedentes del innovador

Básicos

Nombre: Cristián Andrés Rodríguez Arenas

RUT: 15.635.473-2

Ocupación: Profesor Universitario y Consultor

Institución: Pontificia Universidad Católica de Chile (PUC)

Formación

Título Profesional: Ingeniero Civil de Industrias mención Tecnologías de la Información

Institución: Pontificia Universidad Católica de Chile (PUC)

Experiencia relevante para el proyecto

Profesional

1. Asesoría para la implementación de un modelo de Gobierno y Calidad de Datos para AFP Cuprum.
Este modelo fue validado por Principal International y Principal USA como modelo referente para América Latina. Ha estado vigente y madurando por más de 5 años, con crecimiento sostenido y visibilidad a nivel de los gerentes de primera línea de la organización. Es considerado una pieza clave para el proceso de transformación digital.
2. Asesoría para el levantamiento y propuesta de un modelo de gobierno de datos e inteligencia de negocios para la Pontificia Universidad Católica de Chile.
Este modelo incluye los procesos, roles, tecnología y metodologías asociados al gobierno de datos.
3. Asesoría para la implementación de un modelo de Gobierno de Datos en Hogar de Cristo (2019 en adelante).
4. Asesoría para la implementación de un modelo de Gobierno de Datos en Banchile Inversiones (2019 en adelante).

Académica

Pontificia Universidad Católica de Chile (PUC) (2016 – 2020):

- a. Diplomado Inteligencia de Negocios
 - Curso Gestión de Proyectos de Inteligencia de Negocios
 - Curso Tendencias en Inteligencia de Negocios
 - Curso Gobierno de Datos
 - Curso Pensamiento Analítico
- c. Magister en Tecnologías de la Información y Gestión (MTIG)
 - Seminario TI: Inteligencia de Negocios (primer semestre 2020)
 - Seminario TI: Data Science (segundo semestre 2020)

Resumen

La iniciativa tiene como objetivo principal generar un piloto de Open Data en la ACHS, orientado específicamente a la publicación de datos y visualizaciones de accidentabilidad de trayecto. Los objetivos específicos corresponden a validar cada una de las etapas asociadas a una metodología de publicación de datos en forma abierta bajo estándares Open Data, llamada COMSODE. Estas etapas son preparar un plan de publicación de los datos, preparar los datos para su publicación, ejecutar la publicación y definir un plan de archivado y manejo de la historia.

Como resultado se obtiene no sólo la validación de la metodología y sus actividades, que se ve son apropiadas y aplicables a la realidad de una mutual, sino que también se genera un sitio web, con carácter de “piloto”, donde se publican visualizaciones y los datos definidos dentro del proyecto. Los usuarios de esta información pueden acceder a ellos de forma libre y gratuita, sin restricciones de acceso alguno.

Finalmente, se concluye que la factibilidad técnica y operacional no garantiza el éxito de este tipo de iniciativa en el tiempo, si no que requiere de un compromiso de la organización a largo plazo, tanto en la operación como en el involucramiento de alto nivel.

ÍNDICE

Antecedentes del innovador	2
Básicos.....	2
Formación	2
Experiencia relevante para el proyecto	2
Profesional	2
Académica	2
Resumen.....	3
Introducción y antecedentes	6
Introducción	6
Antecedentes	7
Definición del desafío.....	9
Objetivos	9
Definición de los usuarios o dolientes y masa crítica para viabilidad de la solución.....	10
Usuarios.....	10
Masa crítica	11
Revisión de la literatura, o experiencias relevantes	11
Experiencias relevantes.....	11
Transbase	11
Ciudad de Rennes, Francia	12
Portal de datos abiertos de Europa.....	14
Etapas del diseño de la innovación	15
Metodología	15
Metodología aplicada al proyecto.....	16
1. Desarrollar un plan de publicación de los datos de Accidentabilidad de Trayecto bajo principios de Open Data, que considere aspectos técnicos, legales y comunicacionales	16
2. Preparar los datos de Accidentabilidad de Trayecto para ser publicados, considerando el tratamiento de los datos y la habilitación de la plataforma tecnológica asociada.....	17
3. Ejecución de la Publicación de los datos de Accidentabilidad de Trayecto, disponibilizando los datos en un portal web en conjunto con visualizaciones de apoyo	17
4. Definición proceso archivo de la información, para tener en cuenta al momento de decidir dejar de publicar los datos de Accidentabilidad de Trayecto	18
Desarrollo de la innovación.....	19

Fase 1. Desarrollar un plan de publicación de los datos de Accidentabilidad de Trayecto	19
2. Preparar los datos de Accidentabilidad de Trayecto para ser publicados	26
3. Ejecución de la Publicación de los datos de Accidentabilidad de Trayecto	29
4. Definición proceso archivo de la información.....	29
Resultados	30
Sitio Web	30
Visualizaciones	30
Vista general del sitio web	34
Cumplimiento de Objetivos:	34
Conclusiones	35
Anexo: Detalle Metodología COMSODE	37
Anexo: OECD Open Data	37

Introducción y antecedentes

Introducción

Hoy no existe un mecanismo que permita acceso fácil y transparente a datos, de las Mutuales, que son de interés para la comunidad en general (estudiantes, investigadores, innovadores, otros organismos públicos y privados, etc.).

Esto produce una serie de desafíos y oportunidades, entre ellos:

- Genera desconfianza en el público y en otras organizaciones: ¿Cómo sabemos qué hay detrás de los indicadores presentados por la organización? ¿Están bien calculados? ¿La metodología de cálculo es consistente a través de los años?

- Genera ineficiencias para compartir información entre entidades: Hoy las distintas solicitudes de información deben realizarse por canales formales (e informales) que involucran una serie de pasos y mecanismos para la obtención de la información, lo que puede demorar semanas (o meses) y en algunos casos el desfase temporal puede hacer que los datos requeridos o el análisis que se desea hacer pierda valor e incluso deje de tener sentido.

- Dificulta la labor de investigación de organismos y personas: Toda iniciativa de investigación de una universidad, instituto, centro, investigador, etc. depende de su capacidad de gestión e iniciativa para dedicar recursos a obtener acceso a la información, lo cual además de generar un tiempo de espera -que puede ser considerable- muchas veces termina con un acceso parcial a la información solicitada, lo que dificulta la labor investigativa y a veces no permite sacar conclusiones de valor (y veraces) de los estudios realizados.

- Limita la innovación por parte de la comunidad: El acceso público (e irrestricto) a la información ha generado una ola de innovaciones en distintas partes del mundo, por ejemplo:

- o Transbase (<http://transbasesf.org/transbase/>) es una iniciativa del departamento de salud pública de San Francisco (SFDPH) basado en principios de Open Data que, con la colaboración de otros organismos, permite visualizar distintas variables que impactan en la salud de las personas. Comenzó publicando datos de accidentabilidad de peatones a través del concepto de corredores de alta accidentabilidad, esto corresponde al 6% de las millas de las calles donde se generaban el 60% de los accidentes peatonales severos y fatales. Estos mapas generaron campañas de educación, concientización, e incluso, el desarrollo de soluciones viales para reducir los accidentes.

- o Algunas ciudades, como Rennes en Francia, publican de forma abierta toda la información relativa a su presupuesto y gasto real, además de ofrecer visualizaciones para facilitar el acceso a ésta (<http://dataviz.rennesmetropole.fr/budget/>). Esto ha generado una sensación de transparencia que se ha traducido en una recuperación de la confianza en el accionar público por parte de la ciudadanía.

- o UrbanDataAnalytics (<http://www.urbandataanalytics.com/en/>) es un emprendimiento enfocado en ayudar a las personas a tomar mejores decisiones respecto a la compra de inmuebles. Combinan diferentes fuentes de información pública para generar visualizaciones interactivas que permitan analizar de manera más intuitiva y rápida diversos factores y variables que son valiosos para este tipo de decisiones.

Antecedentes

El concepto de los datos abiertos (open data) es una filosofía y práctica que persigue que determinados tipos de datos estén disponibles para la comunidad en general, sin restricciones de derechos de autor, de patentes o de otros mecanismos de control. Sus fundamentos siguen principios similares a los del software libre, siendo una de las más importantes características el libre acceso de los datos.

Esta es una iniciativa global que persigue que datos e información de organismos públicos estén disponibles para ser utilizados, reutilizados y redistribuidos libremente para y por cualquier persona, de manera que puedan ser aprovechados por ciudadanos y empresas.

Incluye todos aquellos datos e información no personal, es decir, datos que no contienen información sobre personas específicas e individualizadas.

Para que un conjunto de datos sea considerado “datos abiertos” debe cumplir con un grupo de principios.

- Completos: Todos los datos públicos se ponen a disposición. Que no están sujetos a limitaciones de privacidad, seguridad o privilegios
- Primarios: Los datos se recogen en la fuente, con el más alto nivel de granularidad, no hay agregaciones o modificaciones
- Oportunos: Los datos se pondrán a disposición tan pronto como sea necesario para preservar el valor de los datos
- Accesibles: Los datos están disponibles para la gama más amplia de usuarios con la más amplia gama de propósitos
- Procesable automáticamente: Los datos se estructuran razonablemente para permitir el procesamiento automatizado
- Acceso no discriminatorio: Los datos están disponibles a cualquier persona, sin necesidad de registro.
- Formatos no propietarios: Los datos están disponibles en un formato sobre el cual ninguna entidad tiene el control exclusivo
- Libres de licencia: Los datos no están sujetos a ningún derecho de autor, patentes, marcas o regulación.

La apertura de los datos trae consigo beneficios tanto para las instituciones públicas como la ciudadanía.

A nivel de instituciones, genera beneficios de interoperabilidad entre servicios públicos, crea más y mejores políticas públicas y potencia la colaboración ciudadana para la mejora de los servicios públicos.

A nivel de la ciudadanía, aumenta la transparencia del sistema y la generación de nuevos servicios con valor social.

Open Data fue propuesto en diciembre del 2007 por un panel de 30 expertos del grupo de trabajo de Gobierno Abierto de Estados Unidos. Actualmente, son los encargados de mantener actualizados los principios y filosofía sobre el que se sustentan. <https://opengovdata.org/>

Definición del desafío

El desafío es crear un piloto de Open Data que abarque la publicación de los datos de Accidentabilidad de Trayecto de la Asociación Chilena de Seguridad.

A través de una metodología internacional de Open Data (COMSODE), se define y construye el proceso de publicación de esta información, cuyo entregable final es un sitio web, administrado por la ACHS para efectos del piloto, que tiene disponible los datos asociados a la Accidentabilidad de Trayecto, además de visualizaciones de Apoyo para el entendimiento de los datos.

Este proyecto busca, además, validar la efectividad de la metodología para la publicación de datos por parte de una Mutual, además del impacto sobre la comunidad de disponer de datos bajo estándares Open Data, lo que permitiría la adopción por parte del resto de las Mutuales. Mientras más información esté disponible, mayor la capacidad de análisis y de obtener conclusiones valiosas a partir de la información.

Objetivos

El objetivo general es desarrollar un piloto de Open Data en la mutualidad enfocado en los datos de Accidentabilidad de Trayecto

Los objetivos específicos son los siguientes:

1. Desarrollar un plan de publicación de los datos de Accidentabilidad de Trayecto bajo principios de Open Data, que considere aspectos técnicos, legales y comunicacionales.
2. Preparar los datos de Accidentabilidad de Trayecto para ser publicados, considerando el tratamiento de los datos y la habilitación de la plataforma tecnológica asociada.
3. Ejecución de la Publicación de los datos de Accidentabilidad de Trayecto, disponibilizando los datos en un portal web en conjunto con visualizaciones de apoyo.
4. Definición del proceso de archivo de la información, para tener en cuenta al momento de decidir dejar de publicar los datos de Accidentabilidad de Trayecto.

Definición de los usuarios o dolientes y masa crítica para viabilidad de la solución

Usuarios

A través de un análisis de la experiencia respecto a Open Data, podemos definir los siguientes usuarios, señalando los problemas y beneficios que obtienen de esta iniciativa:

- Académicos:
 - Dolores:
 - Dificultad para acceder a datos detallados y de calidad para realizar estudios y análisis.
 - Procesos lentos y engorrosos para obtener la información.
 - Beneficios:
 - Acceso libre y sin restricciones a data a gran nivel de detalle.
 - Eficiencia en la adquisición de datos.
- Ciudadanos:
 - Dolores:
 - Falta de confianza en los indicadores y estudios presentados por los organismos públicos.
 - Falta de acceso a datos para realizar análisis.
 - Beneficios:
 - Acceso libre y sin restricciones a data a gran nivel de detalle.
 - Eficiencia en la adquisición de datos.
- Emprendedores:
 - Dolores:
 - Falta de acceso a datos para generar ideas, generar y validar hipótesis, idear nuevos procesos y formas de capturar valor en base a estos datos.
 - Beneficios:
 - Acceso libre y sin restricciones a data a gran nivel de detalle.
 - Eficiencia en la adquisición de datos.
- Servidores públicos:
 - Dolores:
 - Falta de acceso a datos para responder preguntas de interés.
 - Procesos ineficientes de solicitud de información que pueden demorar días, semanas o meses.
 - Beneficios:
 - Acceso libre y sin restricciones a data a gran nivel de detalle.
 - Eficiencia en la adquisición de datos.
- Prensa:
 - Dolores:
 - Falta de acceso a datos para responder preguntas de interés.

- Procesos ineficientes de solicitud de información que pueden demorar días, semanas o meses.
- Dificultad para analizar información en detalle.
- Beneficios:
 - Acceso libre y sin restricciones a data a gran nivel de detalle.
 - Eficiencia en la adquisición de datos.

Masa crítica

Para efectos del piloto, no se define una masa crítica para validar el éxito de la iniciativa. Se desea validar la factibilidad más que el éxito.

Revisión de la literatura, o experiencias relevantes

Experiencias relevantes

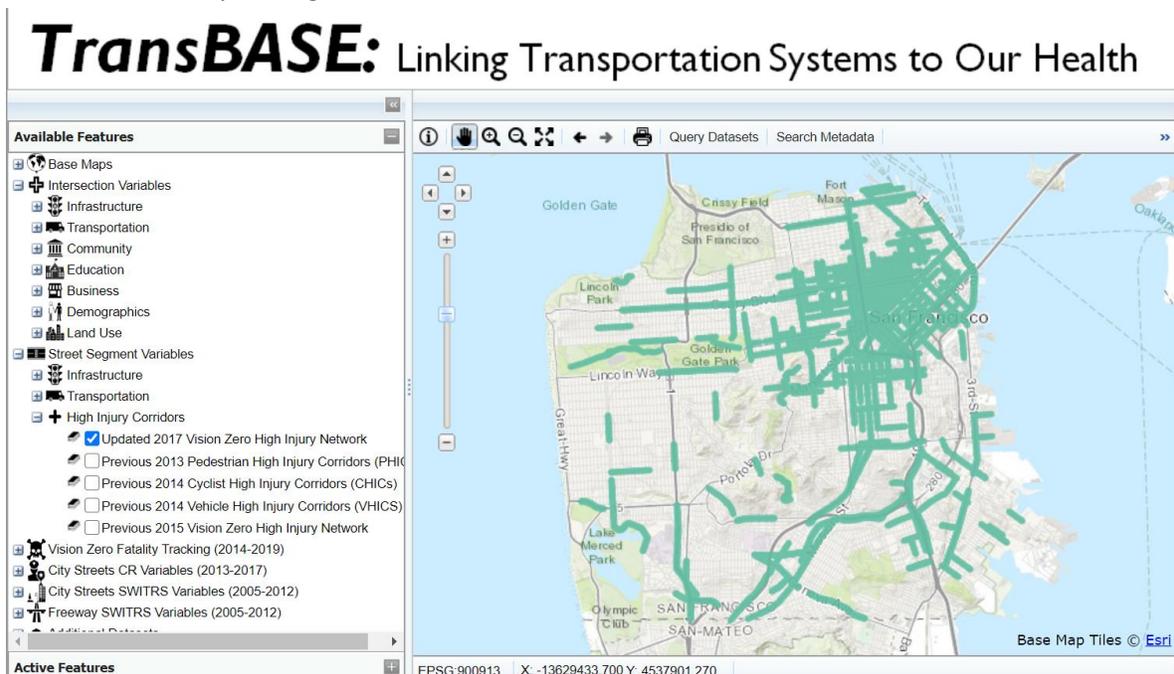
Transbase

Transbase (<https://transbase.sfgov.org/transbase/>), del departamento de Salud Pública de San Francisco, Estados Unidos, presenta un sitio web público orientado a mostrar, a través de un sistema de información geográfica (GIS), distinta información relacionada a la ciudad.

Es un sitio de fácil navegación e interacción dinámica.

A continuación dos ejemplos:

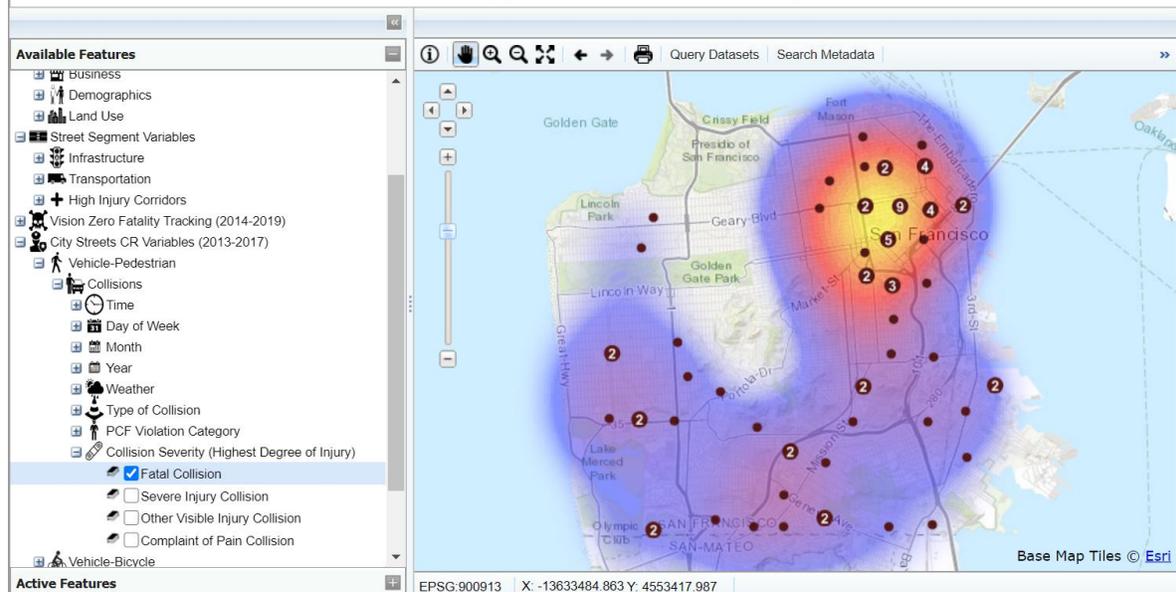
a) Sectores de mayor riesgo de accidentes al 2017



En el mapa se marcan en verde los sectores (calles e intersecciones) donde existe un mayor riesgo de accidentes. Esto permite a la comunidad entender mejor el riesgo de circular en las distintas zonas de la ciudad.

b) Mapa de calor de accidentes fatales (2013 – 2017)

TransBASE: Linking Transportation Systems to Our Health



En el mapa se pueden ver colores cuya intensidad asociado a la cantidad de colisiones fatales en las distintas zona (morado menor cantidad, después rojo y finalmente amarillo).

Ciudad de Rennes, Francia

Esta ciudad en Francia tiene una política bastante madura de Open Data, orientada tanto a dar acceso a datos en forma irrestricta como en entregar visualizaciones de apoyo para la comprensión y análisis de los datos (<http://dataviz.rennesmetropole.fr/>).

A continuación, dos ejemplos:

a) Pago de servicios cofinanciados por la municipalidad

SERVICES AUX USAGERS : QUI PAYE QUOI ?



Le prix payé directement par les usagers des services municipaux représente une part minoritaire du coût réel de ces services.



Este gráfico muestra, para algunos servicios que ofrece la ciudad, cuál es el costo real del servicio y cómo se desglosa entre quienes lo financian. Así, por ejemplo, la primera columna corresponde al servicio de jardín infantil, e indica que del costo real por hora (10 euros), el ciudadano paga sólo 1,2 euros, el resto viene de otras fuentes de financiamiento.

a) Portal de acceso a los datos

La ciudad ofrece además un portal de acceso irrestricto a todos los conjuntos de datos que dispone (350 en la actualidad). Los usuarios pueden buscar y descargar los datos que deseen, y utilizarlos con el fin que estimen convenientes, sin restricciones.

Portal de datos abiertos de Europa

Es un sitio público que recolecta datos abiertos de todos los países de Europa y los consolida en un solo lugar, facilitando la búsqueda y descarga de los mismos. Actualmente contiene 1.110.379 conjuntos de datos, abarcando 36 países y 85 catálogos (<https://www.europeandataportal.eu/en>).

The screenshot shows the European Data Portal interface. At the top is a dark blue navigation bar with a home icon and dropdown menus for COVID-19, Data, Impact & Studies, Training, News & Events, and About. Below this is a main content area with a light blue background. On the left, a 'Search Datasets' section features a map of Europe and 15 category icons: Agriculture, Fisheries, Forestry & Foods; Economy & Finance; Education, Culture & Sport; Energy; Environment; Government & Public Sector; Health; International Issues; Justice, Legal System & Public Safety; Population & Society; Regions & Cities; Science & Technology; and Transport. Below the map are buttons for 'DATA CATALOGUES' and 'ALL DATASETS'. On the right, the 'European Data Portal' title is followed by a descriptive paragraph. Below the text are three large digital-style counters: '36 Countries', '85 Catalogues', and '1110379 Datasets'. A small upward arrow icon is in the bottom right corner.

Etapas del diseño de la innovación

Metodología

A continuación, se presenta la metodología de Open Data usada para efectos del piloto, describiendo sus fases y actividades.

COMSODE (Components Supporting the Open Data Exploration), es una metodología de origen Europeo (www.comsode.eu) utilizada para proyectos de Open Data.

En el anexo “Anexo: Metodología COMSODE” se adjunta el documento oficial sobre la metodología.

En resumen, esta metodología considera 4 fases y 4 prácticas transversales:

Las fases son:

1. Desarrollo del plan de publicación:

El objetivo de esta fase es la identificación y análisis de los datasets candidatos para ser publicados y la generación de un plan de publicación, el cual contiene para cada dataset a ser publicado su nivel de apertura, los roles y responsabilidad asociados al proceso de publicación y gestión, los términos y condiciones asociados al uso de los datos y el roadmap asociado a la publicación.

2. Preparación de la publicación:

El objetivo de esta fase es preparar los datasets para ser publicados de acuerdo a las definiciones de la fase anterior. Asociado a esta fase están la configuración de los accesos a los datos, la creación de la metadata, selección e implementación de la plataforma para la publicación. Se debe considerar también en esta fase la anonimización de los datos y la creación de las licencias de uso respectivas.

3. Realización de la publicación:

El objetivo de esta fase es la publicación de los datos de acuerdo a las definiciones realizadas en las fases anteriores. Se debe probar el proceso de publicación, publicar la data y realizar la mantención de estos y los procesos relacionados.

4. Archivo de la información:

El objetivo de esta fase es manejar la etapa final de la vida de un dataset que puede gatillarse por un cambio de legislación o de las políticas de uso de los datos. Estos datos deben dejar de estar disponibles y generar los procesos necesarios para terminar con la publicación y actualización de la información.

Las actividades transversales se realizan de forma continua durante el desarrollo del proyecto, y son:

- Gestión de Calidad de Datos: El objetivo de esta actividad es asegurar la evaluación y mejoramiento de la calidad de los datos en el tiempo.
- Gestión de Comunicaciones:

El objetivo de esta actividad es informar a los distintos usuarios/clientes la disponibilidad de los distintos conjuntos de datos, la promoción del uso de los mismos y la obtención de retroalimentación de los procesos y datos disponibilizados.

- Gestión de Riesgos:

El objetivo de esta actividad es asegurar la correcta gestión de los riesgos asociados a la publicación de datos bajo los principios de Open Data.

- Gestión de Beneficios:

El objetivo de esta actividad es asegurar la correcta gestión y monitoreo de los beneficios asociados a la publicación de datos bajo los principios de Open Data.

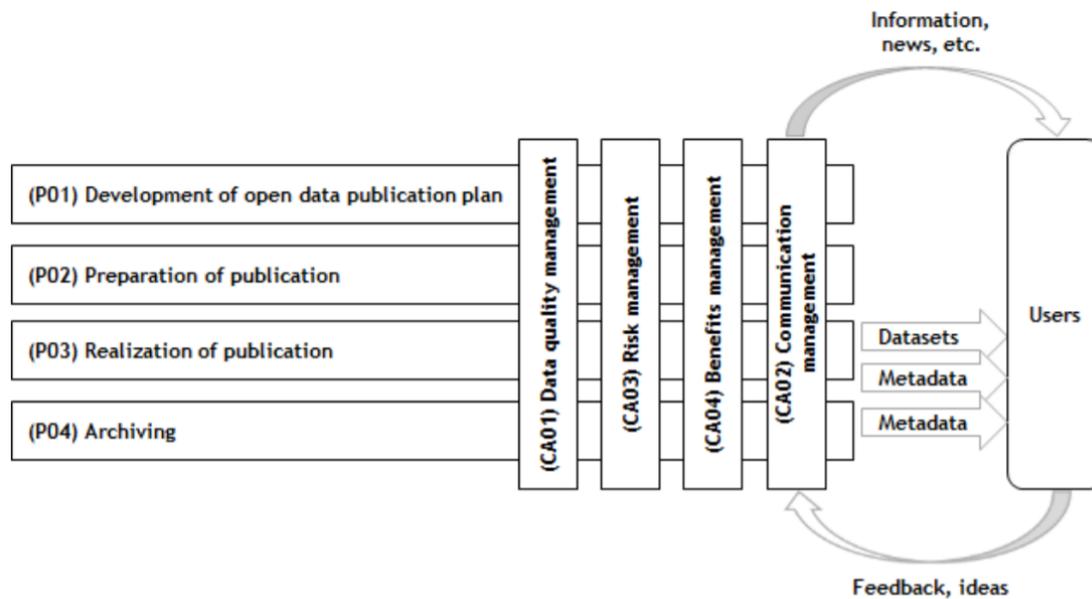


Ilustración 1: Vista general Metodología COMSODE. www.comsode.eu

Metodología aplicada al proyecto

1. Desarrollar un plan de publicación de los datos de Accidentabilidad de Trayecto bajo principios de Open Data, que considere aspectos técnicos, legales y comunicacionales

Actividades

- Análisis de las fuentes de datos. Revisión de las fuentes asociadas a la Accidentabilidad de Trayecto, sus estructuras y datos disponibles para una potencial publicación.
- Determinar nivel de apertura: Definición, en conjunto con las áreas de la ACHS asociadas a Accidentabilidad de Trayecto y datos en general, cuál información debe y no debe publicarse, resguardando la privacidad de las personas y restricciones normativas/legales.
- Identificación Usuarios Potenciales: Análisis de quiénes son los beneficiarios/usuarios de esta información y que valor pueden obtener en base a ella.

- Definición del Plan de Comunicación: Definición de un plan de comunicación orientado a los distintos usuarios potenciales, que permita informar, gestionar expectativas y generar demanda por la información.
- Evaluación de Calidad de Datos: De los datos seleccionados para la publicación se analiza su nivel de calidad de datos, definiendo las acciones correctivas que se deben realizar previo a la publicación.
- Identificación de Riesgos y Beneficios: Análisis de los riesgos asociados al mal uso de la información a publicar, así como los beneficios esperados para los distintos usuarios.

2. Preparar los datos de Accidentabilidad de Trayecto para ser publicados, considerando el tratamiento de los datos y la habilitación de la plataforma tecnológica asociada

Actividades

- Preparación Fuentes de Datos: Gestión de los accesos a las fuentes de datos para la lectura de los datos de interés.
- Definición Esquema de Datos: Definición de la estructura de los datos que serán publicados (metadata) así como el formato de publicación (csv, xls, xlsx, API, etc.).
- Definición Esquema de Integración: Definición arquitectura de la integración, incluyendo métodos (ETL, procedimientos almacenados, API's) y frecuencia de ejecución (diaria, semanal, mensual, etc.).
- Preparación campaña comunicacional: Gestión de los recursos y diseño del material asociado al plan de comunicaciones.
- Diseño e Implementación procesos integración de datos: Diseño y Construcción de los procesos de integración de datos que toman datos desde el origen y los publican.
- Implementación plataforma tecnológica: Diseño y Construcción del sitio web de acceso a la información, así como las visualizaciones de apoyo.
- Pruebas procesos de integración de datos: Pruebas de los procesos de integración validando que se cumplan las restricciones sobre los datos definidas con anterioridad.

3. Ejecución de la Publicación de los datos de Accidentabilidad de Trayecto, disponibilizando los datos en un portal web en conjunto con visualizaciones de apoyo

Actividades

- Publicación Inicial: Disponibilización de los datos a través del sitio web en conjunto con visualizaciones de apoyo.
- Ejecución campaña comunicacional: Lanzamiento campaña comunicacional asociada a los datos publicados.
- Medición nivel de Calidad de Datos: Análisis de calidad de los datos publicados validando que respeten los niveles acordados y restricciones definidas.

4. Definición proceso archivo de la información, para tener en cuenta al momento de decidir dejar de publicar los datos de Accidentabilidad de Trayecto

Actividades

- Análisis necesidades de mantención y almacenamiento de la historia: Definición que mantenciones/actualizaciones se deben hacer, en qué periodos y cuando los datos deben dejar de ser publicados (por ejemplo, por obsolescencia).
- Documentación: Documentación y entrega de Plan de Mantención y otros documentos desarrollados durante el proyecto.

Desarrollo de la innovación

A continuación se describe, en base a las actividades planificadas, el trabajo realizado y resultado de cada una de ellas.

Fase 1. Desarrollar un plan de publicación de los datos de Accidentabilidad de Trayecto

Actividades

Análisis de las fuentes de datos:

En la ACHS, los datos relacionados a la accidentabilidad de trayecto residen actualmente en el ERP (SAP). Además, ACHS realiza una serie de procesos sobre los datos que les permite obtener datos derivados a partir de los datos base.

Asociado a cada siniestro, existen dos tipos de información que se almacenan en la fuente:

1. Datos estructurados:

- Los datos estructurados son aquellos que pueden almacenarse directamente bajo una estructura de tablas y relaciones.
- Aquí se dispone de información relacionada al siniestro (hora, fecha, zona lesionada, etc.), de la persona afectada (Rut, nombre, sexo, edad, etc.), de la empresa del trabajador (Rut, nombre, rubro, etc.), entre otros.

2. Datos no estructurados:

- Los datos no estructurados son aquellos cuyo formato (estructura) impide almacenarlos directamente sobre una estructura de base de datos relacional. Ejemplos de estos son los textos libres, audios, videos, imágenes, etc.
- Aquí se dispone del “relato” del trabajador, que corresponde a un relato verbal que entrega el trabajador sobre el accidente y se escribe como un texto “libre”, sin ninguna estructura.

Además, ACHS procesa estos datos y genera datos derivados, tales como:

1. Georeferenciación:

- A partir de la información del lugar del siniestro (que puede venir escrita de cualquier forma) se realiza un proceso de georeferenciación que permite estimar la latitud y longitud del evento.
- Datos generados: Latitud siniestro, Longitud siniestro

2. Predictores:

- A partir del relato del cliente (texto libre) se realiza un procesamiento del texto que, entre otras cosas, permite estimar (a través de modelos de aprendizaje automático supervisado orientados a la clasificación) ciertos atributos de interés respecto al siniestro.
- Datos generados:

- Predictor Medios de Transporte: El medio de transporte usado por el trabajador al momento del accidente.
- Predictor Tipo de Accidente: El tipo de accidente del siniestro (golpe objetos, caída mismo nivel, etc.).

Tomando esto en consideración, se procede a determinar una lista inicial de datos relevantes para publicar respecto a Accidentabilidad de Trayecto, que se presentan a continuación:

Campo	Descripción
ID Siniestro	Identificador único del siniestro
RutEmpresa	Rut de la Empresa del Trabajador
NombreEmpresa	Nombre de la Empresa del Trabajador
DescripcionRubro	Descripción del Rubro de la Empresa del Trabajador
DescripcionSector	Descripción del Sector de la Empresa del Trabajador
Fatales	Indicador si hubo o no víctimas fatales
Predictor_Medio_Transporte	Medio de transporte probablemente utilizado por el trabajador accidentado al momento del accidente.
Predictor_Tipo_Accidente	Tipo de Accidente probable del trabajador accidentado al momento del accidente.
Fecha_Presentacion	Fecha en formato dd-mm-yyyy en la cual se presentó el accidente
Fecha_Siniestro	Fecha en formato dd-mm-yyyy en la cual ocurrió el accidente
SiniestroHora	Hora en formato hh:mm en que ocurrió del siniestro
Hora	Hora en formato hh en que ocurrió del siniestro
Presentacion_Tardía	Días transcurridos entre la fecha del siniestro y la fecha de presentación
Dia_Semana	Día de la semana en que ocurrió el accidente
Comunas	Comuna donde ocurrió el accidente
Region	Región donde ocurrió el accidente
Rut Trabajador	Rut del trabajador
Nombres	Nombres del trabajador
Apellido Paterno	Apellido paterno del trabajador
Apellido Materno	Apellido materno del trabajado
Sexo	Sexo del trabajador
Edad	Edad del trabajador
Zona_Lesionada	Zona lesionada por parte del trabajador
Tiempo_Perdido	Indicador si hubo o no tiempo perdido (STP - CTP)
Dias_Reposo_Total	Número de días de reposo asociados al accidente hasta la fecha de consulta.
Latitud	Coordenada (latitud) que permite establecer el lugar en el espacio donde ocurrió el accidente
Longitud	Coordenada (longitud) que permite establecer el lugar en el espacio donde ocurrió el accidente

Determinar nivel de apertura:

En esta actividad se procede a analizar cuál es el nivel de apertura que tendrán los datos a publicar considerando las siguientes directrices:

- Granularidad: En la medida de lo posible se deben publicar los datos con el máximo nivel de detalle disponible, con el fin de permitir que los usuarios de éstos tengan máxima libertad en cuanto al análisis que pueden realizar.
- Privacidad: Se debe proteger la privacidad de las personas e instituciones. Cualquier campo, o combinación de campos, que permita identificar a la persona y/o institución afectada debe ser eliminado de los datos a publicar, y se debe asegurar que en publicaciones sucesivas no sean considerados.

Considerando estas dos directrices, se decide eliminar los campos que se listan a continuación, señalando la causa caso a caso:

- Rut de la Empresa: Conociendo la empresa, existe la posibilidad que a través de información disponible en medios de comunicación (prensa, redes sociales, etc.), y usando el resto de los campos de la base, se pueda identificar al trabajador afectado.
- Nombre de la Empresa: Ídem caso anterior.
- Rut del Trabajador: Permite identificar directamente al trabajador afectado.
- Nombres del Trabajador: En conjunto con datos asociados al accidente e información de redes sociales, podría permitir la identificación del trabajador afectado.
- Apellido Paterno del Trabajador: Ídem caso anterior.
- Apellido Materno del Trabajador: Ídem caso anterior.

Otro punto de interés es el caso de los campos derivados, que no provienen directo desde la fuente. En este caso, considerando el valor que generan para efectos del análisis que se puede realizar, se decide incluirlos en la lista de campos a utilizar en el piloto. Queda como tarea pendiente, una vez finalizado el piloto y si se desea continuar, crear “términos de uso y condiciones” en el sitio que aclaren el carácter derivado de estos campos y limite la responsabilidad asociada a la toma de decisiones basado en éstos.

El último campo analizado corresponde al de “ID de Siniestro”, que es un indicador único del siniestro en las bases operacionales. Dado que alguien con acceso, aunque sea parcial, a las bases operacionales podría usar el campo para identificar al trabajador afectado, se decide cambiar este ID de siniestro por un correlativo (1, 2, 3, ..., n) sin relación a las bases operacionales.

Con esto, la lista definitiva de campos a utilizar para el piloto queda definida. Se presentan a continuación, incluyendo para cada campo el origen de datos y dos ejemplos:

Campo	Descripción	Origen	Ejemplo 1	Ejemplo 2
Siniestro	Identificador único del siniestro (no relacionado a las bases operacionales)	BD ACHS	16	44
DescripcionRubro	Descripción del Rubro de la Empresa del Trabajador	BD ACHS	LAVANDERIA Y SERVICIOS DE LAVA	ASERRADEROS, TALLERES DE ACEPI
DescripcionSector	Descripción del Sector de la Empresa del Trabajador	BD ACHS	COMERCIO Y RETAIL	FORESTAL MADERERO
Fatales	Indicador si hubo o no víctimas fatales	BD ACHS	NO	NO
Predictor_Medio_Transporte	Medio de transporte probablemente utilizado por el trabajador accidentado al momento del accidente.	Se obtiene a través de un modelo predictivo que analiza los relatos entregados por los accidentados durante la etapa de admisión	caminar	bicicleta
Predictor_Tipo_Accidente	Tipo de Accidente probable del trabajador accidentado al momento del accidente.	Se obtiene a través de un modelo predictivo que analiza los relatos entregados por los accidentados durante la etapa de admisión	golpe objetos	caída mismo nivel
Fecha_Presentacion	Fecha en formato dd-mm-yyyy en la cuál se presentó el accidente	Proviene de la información que proporciona el accidentado en la etapa de admisión almacenada en las BD de ACHS	25-04-2016	28-11-2016
Fecha_Siniestro	Fecha en formato dd-mm-yyyy en la cuál ocurrió el accidente	Proviene de la información que proporciona el accidentado en la etapa de admisión almacenada en las BD de ACHS	25-04-2016	25-11-2016
SiniestroHora	Hora en formato hh:mm en que ocurrió del siniestro	BD ACHS	14:00	7:45
Hora	Hora en formato hh en que ocurrió del siniestro	Calculo interno BD ACHS	14	7
Presentacion_Tardia	Días transcurridos entre la fecha del siniestro y la fecha de presentación	Calculo interno BD ACHS	0	3
Dia_Semana	Día de la semana en que ocurrió el accidente	Calculo interno BD ACHS	Lunes	Viernes
Comunas	Comuna donde ocurrió el accidente	BD ACHS	santiago	talca
Region	Región donde ocurrió el accidente	BD ACHS	SANTIAGO	MAULE
Sexo	Sexo del trabajador	Proviene de la información que proporciona el accidentado en la etapa de admisión almacenada en las BD de ACHS	Hombre	Hombre
Edad	Edad del trabajador	Proviene de la información que proporciona el accidentado en la etapa de admisión almacenada en las BD de ACHS	65	52
Zona_Lesionada	Zona lesionada por parte del trabajador	Proviene de la información que proporciona el accidentado en la etapa de admisión almacenada en las BD de ACHS	Pies	Ojos
Tiempo_Perdido	Indicador si hubo o no tiempo perdido (STP - CTP)	Proviene de la información asignada por el médico en la atención del paciente.	STP	STP
Dias_Reposo_Total	Número de días de reposo asociados al accidente hasta la fecha de consulta.	Proviene de la información asignada por el médico en la atención del paciente.	0	0
Latitud	Coordenada (latitud) que permite establecer el lugar en el espacio dónde ocurrió el accidente	Se obtiene luego de realizar un proceso de georeferenciación de las direcciones facilitadas por los trabajadores accidentados en la etapa de admisión	-33,4536436	-35,4232444
Longitud	Coordenada (longitud) que permite establecer el lugar en el espacio dónde ocurrió el accidente	Se obtiene luego de realizar un proceso de georeferenciación de las direcciones facilitadas por los trabajadores accidentados en la etapa de admisión	-70,6685866	-71,6484804

Identificación Usuarios Potenciales:

En esta actividad se realiza un análisis de los usuarios potenciales de los datos e información que serán publicados.

Para definir los usuarios potenciales se definen y utilizan los siguientes roles, asociados al uso que se le puede dar a los datos:

- Investigador: Utilizará los datos disponibles para realizar estudios o crear modelos que permitan informar decisiones o entender el impacto y potencial de ciertas variables.
- Consumidor de información: Será un agente “pasivo”, que sólo consumirá la información, probablemente a través de los reportes o visualizaciones, pero no tendrá un rol activo referente a los datos.
- Periodista: Utilizará los datos disponibles para informar al público sobre ciertos aspectos interesantes a partir de los datos. Usará las visualizaciones, pero, además, trabajará con los datos publicados para sacar nuevos análisis y conclusiones. Podrá además combinar estos datos con otros conjuntos de datos complementarios.

- Emprendedor: Su interés en los datos es encontrar oportunidades para desarrollar soluciones o servicios novedosos, tanto con fines sociales como particulares. Podrá además combinar estos datos con otros conjuntos de datos complementarios.
- Servidor público: Persona que trabaja en el mundo público, probablemente asociado a temas de prevención de accidentes, cuyo interés es acceder de forma expedita a la información sin necesidad de solicitudes a través de procesos lentos y engorrosos.

Considerando estos roles se definen los siguientes usuarios potenciales:

- Académicos: Roles investigador, servidor público, emprendedor, consumidor de información.
- Ciudadanos (público general): Roles investigador, emprendedor, consumidor de información.
- Emprendedores: Rol emprendedor.
- Personal del mundo público: Roles servidor público, investigador, emprendedor, consumidor de información.
- Prensa (medios de comunicación): Roles periodista, investigador, consumidor de información.

Definición del plan de Comunicación:

El plan de comunicación hace referencia a como se realizará la difusión, a través de distintos medios, de la iniciativa y piloto de Open Data.

Dado que es un prototipo, se decide no realizar un plan de comunicación como parte del piloto, si en una fase siguiente, cuando el piloto esté validado y se desee continuar.

Aun así, se esboza un plan inicial que basa la difusión de la iniciativa a través de tres canales:

- Página oficial de la ACHS (nota de prensa y enlace)
- Publicación en diferentes medios de prensa
- Publicación y difusión en redes sociales.

Evaluación de Calidad de Datos:

La evaluación de la calidad de los datos consiste en analizar, en base al conjunto de datos definidos para publicación, si cumple con los requisitos mínimos para considerar que la información es adecuada para uso público, y si permite obtener conclusiones de valor o se deben tomar ciertos resguardos.

Para efectos del piloto se revisarán los siguientes atributos de Calidad de Datos (llamados dimensiones de Calidad en la literatura de Gobierno de Datos):

- Completitud: Los datos no presentan valores faltantes, nulos o valores por defecto que no corresponden al dominio del campo.
- Unicidad: Los datos no están duplicados.
- Veracidad: Los datos representan fielmente la realidad.
- Oportunidad: Los datos están actualizados.

Para realizar el análisis se trabaja sobre una base de Accidentabilidad que contiene entre los años 2016 y 2019. Esta base fue construida por el equipo de Analytics de la ACHS en base a los campos solicitados. Esta base contiene 136.373 registros.

El análisis preliminar (en % de cumplimiento de cada dimensión), se presenta a continuación:

Campo	Unicidad (%)	Completitud(%)	Oportunidad (%)	Veracidad (%)
SiniestroHora	100%	100%	100%	100%
DescripcionRubro	N/A	100%	100%	100%
DescripcionSector	N/A	100%	100%	100%
Fatales	N/A	100%	100%	100%
Predictor_Medio_Transporte	N/A	100%	100%	100%
Predictor_Tipo_Accidente	N/A	100%	100%	100%
Fecha_Presentacion	N/A	100%	100%	100%
Fecha_Siniestro	N/A	100%	100%	99,90%
SiniestroHora	N/A	100%	100%	100%
Hora	N/A	100%	100%	100%
Presentacion_Tardia	N/A	100%	100%	99,80%
Dia_Semana	N/A	100%	100%	100%
Comunas	N/A	90%	100%	100%
Region	N/A	90%	100%	100%
Sexo	N/A	100%	100%	100%
Edad	N/A	100%	100%	99,99%
Zona_Lesionada	N/A	99,8%	100%	100%
Tiempo_Perdido	N/A	100%	100%	100%
Dias_Reposo_Total	N/A	100%	100%	99,6%
Latitud	N/A	90%	100%	100%
Longitud	N/A	90%	100%	100%

En el caso de Unicidad, sólo nos interesa que el correlativo (asociado al ID de siniestro) sea único. Se cumple.

En el caso de completitud, vemos dos casos:

- Para los campos asociados a la georreferenciación (Región, Comunas, Latitud, Longitud) existe un 10% de registros donde ésta no funciona, y los campos quedan en blanco.
- Para el caso de Zona Lesionada, aparece un 0,2% de los registros con el valor “NULL”.

Respecto a la oportunidad, los registros se actualizan mensualmente en el ERP de forma automática, no hay procesamiento externo de la información.

Finalmente, respecto a la Veracidad, tenemos varios casos:

Los campos en celeste, unos asociados a la georreferenciación y otros estimados por algoritmos en base al relato, la veracidad depende del grado de confianza en el resultado del proceso relacionado. Para nuestros efectos consideramos que estos procesos están validados y certificados, por tanto, podemos creer en la veracidad de sus resultados.

- En el caso de la fecha de siniestro, se consideran veraces los registros con fecha superior a septiembre del 2015, dando un 0,1% de los registros que no cumplen. Muy bajo.
- En el caso de Presentación Tardía, que se mide en días, se consideran veraces los registros cuyo valor es distinto a -1 y menor o igual a un año. Existe un 0,2% de los registros que no cumplen. Muy bajo.
- En el caso de la Edad, se consideran veraces los registros con edad mayor o igual a 18 años. Existe un 0,01% de los registros que no cumplen. Muy bajo.
- En el caso de los días de reposo, se consideran veraces los días de reposo menores o iguales a 365. Existe un 0,4% de los registros que no cumplen. Muy bajo.

Se debe señalar que las reglas de calidad definidas se basan en un análisis de los valores que toman los campos en la base (proceso llamado perfilamiento de datos en la literatura de Gobierno y Calidad de Datos) y conversaciones con ACHS, para determinar lógicas razonables y sencillas de implementar.

Evaluar la veracidad de los registros requiere la implantación de una metodología de gobierno de datos que involucre iniciativas de calidad.

Identificación de Riesgos y Beneficios:

En base a un análisis de riesgos se genera la siguiente matriz, que presenta los riesgos identificados, su probabilidad, impacto y plan de mitigación:

ID	Factor de Riesgo	Tipo de Riesgo	Probabilidad	Impacto	Mitigación
1	Hackeo del sitio	Técnico	Bajo	Bajo	Backup del sitio
2	Baja calidad de los datos	Técnico	Bajo	Bajo	Validación datos antes de publicación
3	Virus y Malware	Técnico	Media	Medio	Backup del sitio
4	Violación propiedad intelectual	No-Técnico	Bajo	Alto	Validar contenido antes de publicar con área legal
5	Alteración de la data	No-Técnico	Bajo	Alto	Validación datos antes de publicación

Respecto a los beneficios, también se realiza un análisis de beneficios que genera la siguiente matriz, que presenta los beneficios, quien percibe el beneficio y su descripción:

ID	Beneficio Potencial	Horizonte	Beneficio Para	Descripción
1	Mejora calidad de datos	Corto	ACHS	Proceso de revisión permite mejorar la calidad de los datos
2	Mayor transparencia	Corto	ACHS	Datos abiertos entrega mayor confianza y transparencia para ACHS
3	Nuevas iniciativas	Mediano	ACHS	Interesados se acercan a ACHS para realizar nuevas iniciativas
4	Mayor investigación	Mediano	Usuarios	Acceso eficiente a datos abiertos permite mayor y mejor investigación
5	Datos de apoyo	Mediano	Usuarios	Usuarios disponen de datos de apoyo para análisis e investigaciones
6	Resolución de problemas	Largo	Usuarios	Fomenta creatividad y resolución de problemas existentes
7	Mayor participación ciudadana	Largo	Sociedad	Mejor acceso a datos de calidad que informan decisiones
8	Innovación	Largo	Sociedad	Junto con otros conjuntos de datos abre la puerta a mejores servicios públicos

2. Preparar los datos de Accidentabilidad de Trayecto para ser publicados

Actividades

Preparación Fuentes de Datos:

Al analizar las fuentes de datos, queda claro que los datos ya están disponibles para ser consumidos por el proceso de Datos Abiertos, por lo que la preparación de las fuentes no es necesaria.

Se evalúa el impacto de los procesos de extracción de datos, pero considerando que la información se manejará mensualmente y el conjunto de datos no es muy extenso, se considera bajo.

Definición Esquema de Datos:

Existen dos grandes formas de compartir la información:

- Acceso a la data subyacente:
Consiste en dar la posibilidad a los usuarios que accedan a los datos, con la mayor granularidad (detalle) posible a través de la plataforma, con algún mecanismo de acceso.
- Acceso a visualizaciones basados en datos abiertos:
Consiste en dar la posibilidad a los usuarios de acceder a distintas visualizaciones, en un sitio público, para que puedan navegar y analizar la información. El objetivo de estas visualizaciones es presentar información de interés para los usuarios y permitir que estos se familiaricen con los datos, su significado y usos.

En el piloto se decide incorporar ambas formas de compartir información.

Respecto a la data subyacente, se realizan las siguientes definiciones:

1. Esquema del modelo de datos:

Para el esquema del modelo de datos, se consideran los siguientes factores:

- a. El volumen de datos no es muy grande (136.000 registros en 3 años)
- b. El crecimiento anual estimado es pequeño (aprox. 40.000 manual)
- c. Los usuarios, en general, no son expertos en diseño de base de datos
- d. Los usuarios tienen familiaridad con Excel
- e. Los usuarios están acostumbrados, en general, a trabajar con “sábanas” de datos, que contienen toda la información en una gran tabla.

Considerando todos estos factores, se decide compartir los datos en formato csv (Comma Separated Values), usando como modelo una “sábana” de datos asociada a los campos definidos para publicar. El proceso de integración de datos se encargará de la transformación de los datos para construir esta “sábana”.

2. Forma de compartir los datos:

Los datos se compartirán a través de un archivo csv disponible a través del portal de las visualizaciones.

3. Formato del archivo:

El formato será csv, dado que es agnóstico a una plataforma o solución particular, y puede ser consumido por diferentes herramientas de forma eficiente.

Definición Esquema de Integración

La figura a continuación presenta el esquema de integración definido:

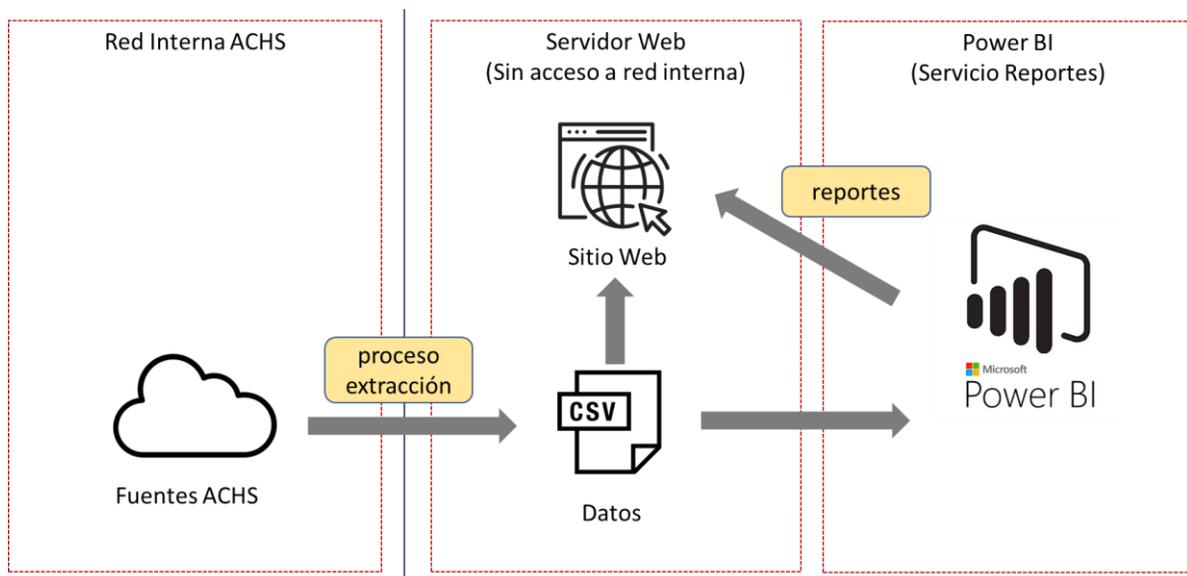


Ilustración 2: Esquema integración

El esquema funciona de la siguiente manera:

1. Los datos residen en las fuentes de la ACHS.
2. Un proceso de extracción toma los datos desde las fuentes ACHS y los lleva a un servidor web, que no tiene acceso a la red interna de la ACHS. Los datos llegan como un archivo separado por comas (csv) de acuerdo con la estructura definida. La noción detrás de esta separación de ambientes es que, en caso de una vulneración de seguridad (hackeo), sólo se pueda acceder a lo que está en el archivo o en la web, que por su naturaleza es data abierta.
3. En el servidor los datos actualizados se ponen a disposición a través del sitio web de datos abiertos.

4. Se utiliza un servicio de reportes web, Power BI de Microsoft, que permite crear, publicar y embeber reportes. Además, cuando se definen los reportes como públicos, compartir y embeber éstos no tiene costos.
5. Los reportes generados en Power BI son embebidos al interior del sitio web de datos abiertos.
6. El sitio web contiene tanto los reportes (visualizaciones) como los datos en formato csv.

Respecto a la selección de Power BI como herramienta de visualización, se hace hincapié en lo siguiente:

- Existen otras herramientas que ofrecen la misma funcionalidad y también en forma gratuita, pero la ACHS está realizando una importante inversión en la plataforma de Microsoft, por lo cual existe actualmente el conocimiento interno para administrar y dar continuidad a los reportes.
- Existe otra alternativa que es crear las visualizaciones a través de algún lenguaje de programación (R, Python, etc.), pero se descarta producto que, a pesar de que el conocimiento existe, la productividad en la creación y administración de reportes es más baja.
- Power BI es una solución en la nube que cuenta con todo el respaldo de Microsoft, incorporando un nivel de seguridad muy alto, que dificulta ataques informáticos que busquen alterar los reportes publicados.
- Considerando lo anterior, es importante señalar que la variedad y cantidad de visualizaciones disponibles en Power BI no es grande, por lo que existe complementariedad a través de:
 - a. Usar Python o R para desarrollar nuevas visualizaciones al interior de Power BI
 - b. Desarrollar nuevas visualizaciones en cualquier lenguaje/programa y embeberlas en el sitio, en conjunto con las de Power BI.

Preparación campaña comunicacional

Dado que no se realizará una campaña comunicacional, esta etapa no aplica.

Diseño e Implementación procesos integración de datos

El proceso de integración es desarrollado por la ACHS, quienes tienen el conocimiento de las fuentes y el equipo con los conocimientos necesarios para construirlo. El producto final del proceso es el archivo señalado anteriormente.

Implementación plataforma tecnológica

Se implementa la plataforma tecnológica en los distintos niveles:

- Sitio web:
El sitio se desarrolla usando HTML5, CSS y Javascript, sin un framework particular. Para el diseño, se usa como base el sitio de la ACHS, con sus estilos e imágenes características.
- Reportes Power BI:
Los reportes son diseñados en conjunto con la ACHS, con activa participación de la contraparte.

Los resultados de la implementación se verán en la sección de Resultados.

Pruebas procesos de integración de datos

Se realizan pruebas del proceso de integración y, dado que el proceso es simple, los resultados son exitosos.

3. Ejecución de la Publicación de los datos de Accidentabilidad de Trayecto

Actividades

Publicación Inicial

La publicación se realiza en calidad de piloto, el sitio web temporal está disponible en la siguiente URL:

<http://54.175.86.96/achs/home.html>

Ejecución campaña comunicacional

No aplica

Medición nivel de Calidad de Datos

Este monitoreo se debe hacer durante algunos meses luego de la publicación y la finalización del piloto, por parte del equipo de ACHS.

4. Definición proceso archivo de la información

Actividades

Análisis necesidades de mantención de la historia:

Para el caso del piloto, se levantan los siguientes puntos:

1. La información se actualizará, en la medida de lo posible, con periodicidad mensual.
2. Esta actualización agregará información al archivo, sin reemplazarla.
3. En el caso que se detecte la necesidad de hacer cambios en la historia, se reprocesará la historia y se volverá a publicar el archivo.
4. Dado que el volumen de datos es bajo, no es necesario archivar o eliminar historia en el tiempo. Se adopta una estrategia de mantener toda la historia en el mediano plazo, comenzando desde el 2016.

Documentación

Para efectos de la documentación, se entregan los informes de avance y el informe final.

Resultados

Sitio Web

Actualmente el sitio web está disponible en la siguiente URL: <http://54.175.86.96/achs/home.html>

Visualizaciones

A continuación, se presentan las visualizaciones desarrolladas con la descripción de cada una:

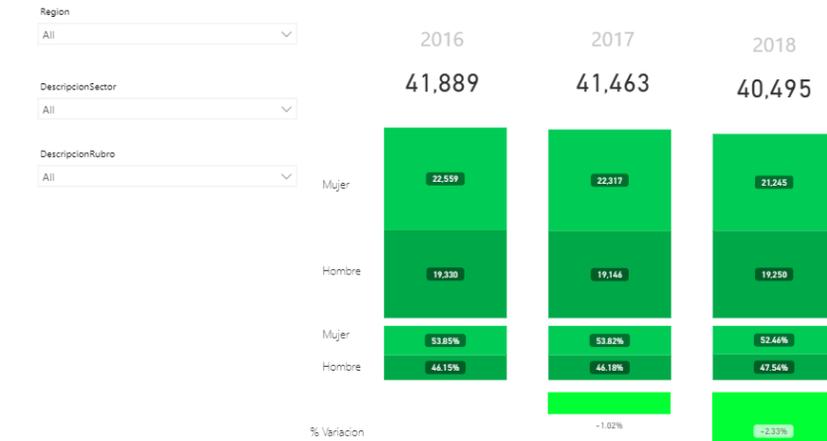
a. Vista general de accidentes por año y sexo (cantidad, % y crecimiento)

Análisis por Año

En este espacio puedes conocer más de la evolución de los accidentes de trayecto de ACHS en los últimos 3 años.

Además, puedes ver cuantos de estos accidentes corresponden a mujeres y hombres (en cantidad y porcentaje).

Te invitamos a usar los selectores del lado izquierdo para profundizar la mirada a través de regiones, sectores y rubros industriales.



b. Vista de análisis por Región de accidentes por año, sexo y edad (interactivo)

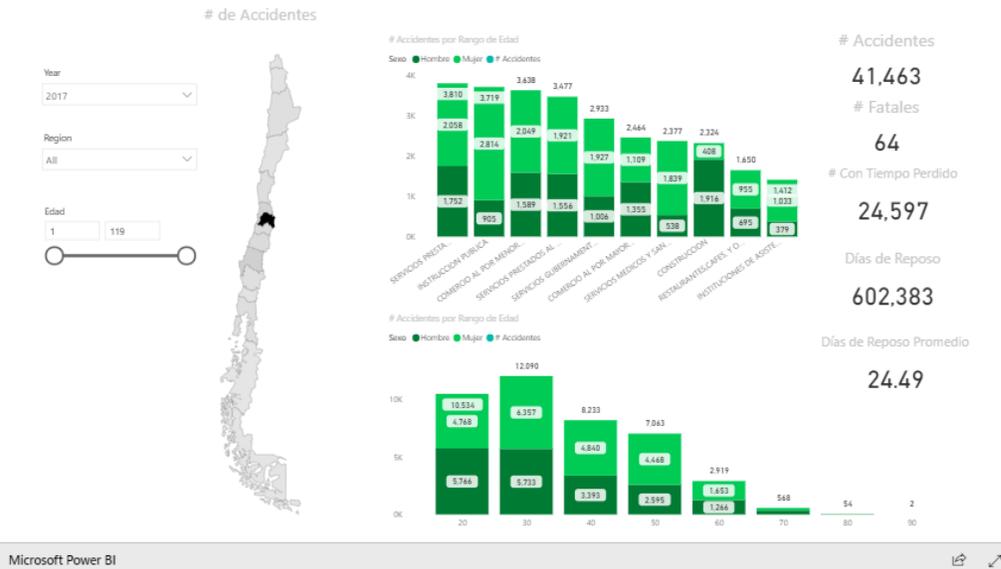
Análisis por Región

En este espacio puedes conocer más respecto a la cantidad de accidentes en el territorio nacional para un año particular.

Además, puedes ver el desglose por rubro/sexo y edad/sexo, junto con otros indicadores generales.

Te invitamos a usar los selectores del lado izquierdo para profundizar la mirada a través de año, región y rango de edad.

¡El mapa es interactivo! Puedes pinchar (hacer click) en una región para filtrar por esta.



c. Vista de análisis comparativo por Región (últimos dos años por sexo y con crecimiento)

Análisis por Sexo

En este espacio puedes conocer más respecto a la cantidad de accidentes por sexo para los últimos dos años, con el desglose por región.

Para cada año puedes ver la distribución por sexo en cantidad y en porcentaje

Distribución de Accidentes por Sexo en Regiones



d. Evolución Mensual a través de los años

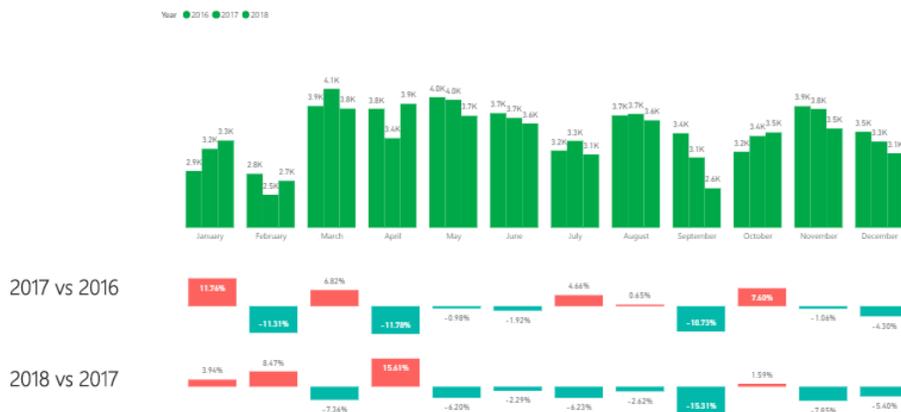
Evolución Mensual Comparada

Un análisis clásico de la accidentabilidad de trayecto corresponde a la presencia de estacionalidades mensuales en su comportamiento.

En el siguiente gráfico, puedes ver como se compara cada mes en los tres últimos años, además de la comparación entre años.

Es interesante notar que en general la accidentabilidad de trayecto a ido descendiendo en los últimos años.

Evolución Mensual # Accidentes a través de los años



Microsoft Power BI

e. Vista análisis de accidentes por tipo y medio de transporte

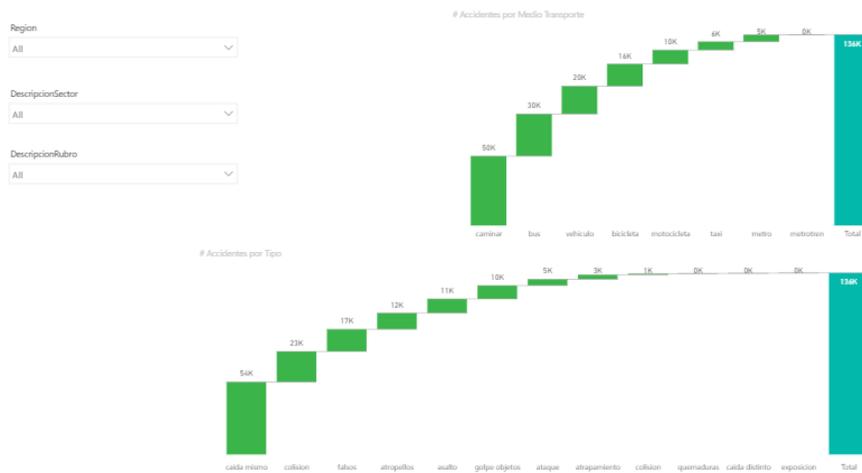
Análisis por Medios de Transporte

Es importante conocer cuáles son los medios de transporte y tipo de accidente más comunes en la accidentabilidad de trayecto.

En el siguiente gráfico, puedes ver de forma ordenada (los primeros tienen más accidentes de trayecto) el desglose por estos dos conceptos.

Es interesante notar que el medio de transporte que más accidentes tiene relacionados es caminar y la principal causa de accidentes es caída del mismo nivel.

Te invitamos a usar los selectores del lado izquierdo para profundizar la mirada a través de región, sector y rubro industrial.



Microsoft Power BI

Vista general del sitio web

A continuación, una visión general del piloto temporal de sitio web:



Análisis por Año

En este espacio puedes conocer más de la evolución de los accidentes de trayecto de ACHS en los últimos 3 años.

Además, puedes ver cuantos de estos accidentes corresponden a mujeres y hombres (en cantidad y porcentaje).

Te invitamos a usar los selectores del lado izquierdo para profundizar la mirada a través de regiones, sectores y rubros industriales.

Cumplimiento de Objetivos:

A continuación, veremos uno a uno los objetivos del piloto:

1. **Desarrollar un plan de publicación de los datos de Accidentabilidad de Trayecto bajo principios de Open Data, que considere aspectos técnicos, legales y comunicacionales.**

Objetivo cumplido dentro del contexto del piloto. Respecto a la metodología formal, faltó el plan de comunicación.

2. **Preparar los datos de Accidentabilidad de Trayecto para ser publicados, considerando el tratamiento de los datos y la habilitación de la plataforma tecnológica asociada.**

Objetivo cumplido dentro del contexto del piloto.

3. **Ejecución de la Publicación de los datos de Accidentabilidad de Trayecto, disponibilizando los datos en un portal web en conjunto con visualizaciones de apoyo.**

Objetivo cumplido dentro del contexto del piloto. Respecto a la metodología formal, faltó el plan de comunicación.

4. **Definición del proceso de archivo de la información, para tener en cuenta al momento de decidir dejar de publicar los datos de Accidentabilidad de Trayecto.**

Objetivo cumplido dentro del contexto del piloto.

Conclusiones

En base a la experiencia obtenida de aplicación de una metodología de Open Data para publicar datos de accidentabilidad de proyecto en una mutual, podemos entregar las siguientes conclusiones:

- Factibilidad Técnica: Dado el carácter de este tipo de iniciativa, no se requiere gran inversión en recursos técnicos, y sólo se requieren plataformas y tecnologías maduras, disponible en la mayoría de las organizaciones.

- Metodología COMSODE: Esta metodología fue utilizada exitosamente en el piloto, mostrando que es factible y recomendable usarla en iniciativas de Open Data en la realidad local.

- Cultura organizacional: El éxito, en el largo plazo, depende que la organización valore y promueva la cultura de datos abiertos en su interior. Además de considerar los datos como un activo estratégico (cosa que actualmente no realizan muchas organizaciones), debe haber una búsqueda continua de nuevos conjuntos de datos de interés público que pueden ser publicados potencialmente generando valor a los distintos usuarios, tanto dentro como fuera de la organización. Sin un cambio cultural es difícil pensar en la sostenibilidad de la práctica de datos abiertos en el mediano plazo.

- “Miedo” a la publicación: Un tema recurrente, al conversar sobre datos abiertos con los miembros de la organización, es el miedo a que, producto de la publicación de datos detallados y abiertos, usuarios de éstos puedan utilizarlo como un “arma” contra la propia organización, a través de distintos mecanismos:

- Detección de errores en indicadores o estudios publicados
- Detección de situaciones anómalas que no hayan sido vistas por la organización
- Publicación de estudios o análisis que perjudiquen a la organización

A pesar de que estos puntos puedan tener sentido, cabe destacar que la promoción de una cultura de datos abiertos va asociada a una cultura orientada al fracaso y la experimentación, donde los errores son reconocidos como oportunidades de aprendizaje, como un momento de reconocer el error y mejorar. En iniciativas de Open Data reconocer errores e incluso aceptar sugerencias/estudios del público genera mayor confianza y marca para la organización.

- Compromiso Ejecutivo: Las iniciativas de datos abiertos deben contar con el compromiso del más alto nivel de la organización, los cambios culturales señalados anteriormente (asociados tanto a una cultura tolerante a errores y al uso de datos como activo estratégico) deben ser liderados, evangelizados y ejemplificados por los más altos ejecutivos de la compañía. Si esto no ocurre, es difícil asegurar la sostenibilidad en el tiempo de la práctica.

- Participación de la industria: Finalmente, es importante señalar que producto de los efectos de economía de red, donde cada actor adicional agrega un valor marginal mayor que el anterior, produciendo un crecimiento de valor exponencial, la inclusión del resto de las mutuales permitiría un impulso importante a la iniciativa de datos abiertos.

Esto permitiría, además de contar con la totalidad de los datos de la industria, y no sólo una muestra, mostrar un compromiso conjunto por parte de la industria en trabajar en conjunto por la transparencia, confianza, colaboración e innovación del sistema en su conjunto.

Anexo: Detalle Metodología COMSODE

En el siguiente documento pueden acceder a la versión oficial de la metodología. Está en inglés.



Adobe Acrobat
Document

Anexo: OECD Open Data

En el siguiente documento pueden ver un “working paper” de la OECD que profundiza bastante en la temática de Datos Abiertos.



Adobe Acrobat
Document