

PORTAL DE MICRODATOS

¿Qué es?

El portal de microdatos es un portal donde se dispone la información estructurada que ha llegado a la Comisión de la Verdad; allí se organizan los conjuntos de datos (bases de datos o dataset) que han sido aportados por diferentes organizaciones del país (privadas, públicas y organizaciones sociales). Este portal ha sido implementado a través del software de código abierto CKAN, por tanto, puede ser considerado como un repositorio de datos abiertos. Toda la información contenida en el portal de microdatos es información estructurada; es decir, todo lo que se ha migrado al portal de microdatos es información estadística y/o bases de datos (anonimizadas -públicas- o sin anonimizar -privadas-) en función de diferentes temáticas. Las entidades que más información ha aportado son: Agencia Nacional de Tierras, Fiscalía General de la Nación, Unidad de Restitución de Tierras y la Unidad para las Víctimas.

¿Cuál es su historia?

El portal de microdatos surge como una apuesta para dejarle a la ciudadanía un repositorio robusto que centralice las bases de datos sobre violencia y violación de derechos humanos que existen en el país. Adicionalmente, el portal se construye con el fin de entregar una fuente de consulta que permita corroborar lo establecido en el Informe Final, documento que cita varias de las bases de datos resguardadas en el repositorio.

La escogencia de CKAN apunta a uno de los principios de la Comisión de la Verdad: utilizar herramientas de código abierto para reducir los costos de licenciamiento que un software propietario puede conllevar para este tipo de proyectos. Ahora bien, en un principio se usó la versión CKAN Andino, que es una mejora realizada por el gobierno de Argentina al software CKAN; no obstante, el tema de la seguridad y la carga de dataset de grandes tamaños se volvieron limitantes importantes dentro de esta versión, por lo que se optó por adaptar otra rama del CKAN para dar respuesta a estos y otros requerimientos puntuales de la Comisión de la Verdad.

La más utilizada, a pesar de todo, es CKAN, porque es una herramienta que no va a requerir un gasto grande en licencias. De hecho, la que teníamos antes de pasar a esta versión se llamaba CKAN Andino; fue, justamente, una modificación que hicieron en el gobierno de Argentina para el portal de datos del gobierno de Argentina... a lo que voy es a la importancia que tiene CKAN en el mundo, que es

una de las herramientas más utilizadas para la publicación de datos; y sigue siendo la más utilizada... ya existen otras de pago, pero muy rara vez vas a encontrar una persona o empresa que adquiera una herramienta de pago para publicar datos...

Como repositorio que centraliza las bases de datos sobre violencia y violación de derechos humanos del país, CKAN es el primer catálogo de microdatos sobre el conflicto armado colombiano.

Funcionalidad

Usuarios

El portal de microdatos funciona para tres tipos de usuarios:

- **Funcionario-Administrador:** perfil encargado de administrar el portal de microdatos. Tiene funciones de cargar, catalogar, editar metadatos y eliminar dataset. A través de este perfil se crean los grupos de usuarios o miembros para acceder a dataset privados o que tienen permisos especiales.
- **Usuario-Consulta:** personas que pueden navegar por el repositorio y acceder a bases de datos con información estadística o información anonimizada. Dentro de este perfil se encuentran los Usuario-Consulta temporales, que son perfiles habilitados por un mes para la consulta de bases de datos privadas.
- **Funcionario-Editor:** perfiles que pueden cargar dataset y editar metadatos (catalogación). No puede realizar administración de usuarios, y la eliminación de dataset solo se hace de manera lógica (el administrador decide si se hace de manera persistente).

Tareas que el sistema proporciona

Las bases de datos pueden consultarse a través de las siguientes opciones

1) Conjuntos de datos.

Este proceso reúne todas las bases de datos, privadas y públicas, acopiadas por la Comisión de la Verdad dentro del portal de microdatos. Las bases de datos públicas pueden ser accedidas desde cualquier usuario; las bases de datos privadas, requieren permisos especiales para ser consultadas.

Conjuntos de datos

Organizaciones

Grupos

Acerca de

Búsqueda

Conjuntos de datos

Organizaciones

Departamento Admini... 29

Unidad para la Aten... 28

Fiscalía General de... 20

Agencia Nacional de... 13

Centro de Estudios ... 13

Unidad Administrati... 12

Departamento Admini... 11

Fundación Comité de... 11

Banco de la República 10

Rutas del conflicto 10

Mostrar Más Organizaciones

Grupos

Buscar conjuntos de datos...

453 conjuntos de datos encontrados

Ordenar por: Relevancia

Arquidiócesis, diócesis y vicariatos apostólicos del país (base original)

Base de datos de las 77 arquidiócesis, diócesis y vicariatos apostólicos, del país entregada por la Conferencia Episcopal de Colombia. Nota: la referencia al Obispado castrense...

[XLSX](#)

Proyecto Generación de Ingresos y Desarrollo de Capacidades Productivas (anon...)

El proyecto plantea procesos relacionados con la aplicación, sostenibilidad e implementación de modelos de Desarrollo Rural Integral con Enfoque Territorial, la innovación,...

[XLSX](#)

Proyecto Capacidades Empresariales, Confianza y Oportunidad (anonimizada)

Base anonimizada de beneficiarios del proyecto Capacidades empresariales que busca mejorar las condiciones de vida, ingresos, y empleo de la población en condición de pobreza en...

Panel para la visualización de los conjuntos de datos

2) Organizaciones.

Este proceso acopia las organizaciones que han aportado sus bases de datos a la Comisión de la Verdad. Cada organización tiene una o varias bases de datos asociadas. Todos los dataset están asociados a una organización.

Organizaciones

Conjuntos de datos

Organizaciones

Grupos

Acerca de

Búsqueda

Organizaciones

¿Qué son las Organizaciones?

Las organizaciones en CKAN son usadas para crear, gestionar y publicar colecciones de conjuntos de datos. Los usuarios pueden tener diferentes perfiles en una organización, dependiente de su nivel de autorización para crear, editar y publicar

Buscar organizaciones

136 organizaciones encontradas

Ordenar por: Nombre Ascendente

ACDI/VOCA
2 Conjuntos de Datos

Aeronáutica Civil
Unidad Administrativa Especial - AEROCIVIL
5 Conjuntos de Datos

Agencia de Desarrollo Rural - ADR
4 Conjuntos de Datos

Agencia de Renovación de Territorio - ART
2 Conjuntos de Datos

Agencia Nacional de Hidrocarburos- ANH

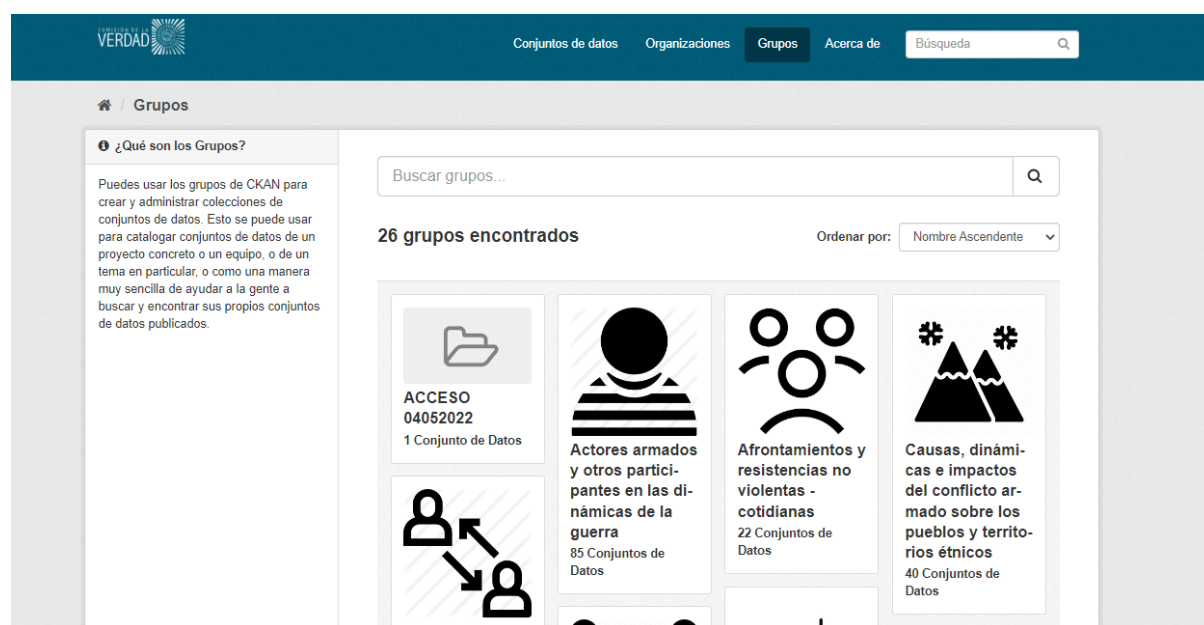
Agencia Nacional de

Agencia Nacional de

Panel para visualización de las organizaciones

3) Grupos.

Este proceso agrupa bases de datos de acuerdo con temáticas. A través de los grupos temáticos, se administran los perfiles temporales con permisos especiales para acceder a las bases de datos privadas contenidas en cada uno de los grupos. Los grupos responden a núcleos o líneas de investigación; en este sentido, cuando se carga una base de datos, se clasifica de acuerdo con temáticas particulares.



Panel para visualizar el conjunto de colecciones o grupos

i El portal de microdatos realiza una trazabilidad de los procesos de carga, modificación, consulta y eliminación de dataset (funcionalidades CRUD). Dicha trazabilidad es visible a través de la opción Flujo de Actividad, que se encuentra en Conjunto de datos y en Organizaciones.

i El portal de microdatos permite realizar búsquedas aplicando diferentes filtros: por organización, por grupos, por etiquetas, por formatos y por licencias.

Requerimientos

Limitaciones y alcances

- El manejo del portal de microdatos es bastante sencillo. Se selecciona la base de datos que se quiere visualizar, la cual se encuentra asociada a alguna entidad, y se ingresa al recurso para explorarlo y/o descargarlo; adicionalmente, es posible identificar cuándo y quién cargó la información (la trazabilidad de las operaciones). La visualización preliminar de los datos se hace a través de un formato estandarizado (Excel), con el fin de que puedan ser consultados por el mayor número de personas; no obstante, el portal recibe varios formatos en función de lo que puede componer un dataset: diccionario, presentaciones en Power Point, scripts en R, etc. Esto implica que un dataset debe tener, como mínimo, un Excel (o un TXT o CSV para bases de datos de varios millones de registros) en tanto es el formato que permite estructurar y establecer una visualización de los datos, tal y como se muestra en el siguiente ejemplo:

Agregar Filtro

Tabla Gráfico Mapa 63 records « 1 – 63 » 🔍 Search data ... Go » Filtros

_id	Subpoblación - Hecho que motiva la solicitud.	AÑO 2018	AÑO 2019	AÑO 2020	-Total general
54	VIOLACIÓN A LOS DDHH E INFRACCIÓN AL DIH	1	0	0	1
13	TORTURA	0	1	0	1
60	RIESGO / VULNERABILIDAD	1	0	0	1
40	DESAPARICIÓN FORZADA	0	1	0	1
45	SECUESTRO	1	0	0	1
2	ABANDONO O DESPOJO FORZADO DE TIERRAS	0	1	0	1
36	SECUESTRO	0	0	1	1
32	DELITOS CONTRA LA LIBERTAD E INTEGRIDAD SEXUAL	0	1	0	1
48	ABANDONO O DESPOJO FORZADO DE TIERRAS	1	0	0	1
51	HOSTIGAMIENTO	1	0	0	1
58	DESPLAZAMIENTO FORZADO	2	0	0	2
6	DESAPARICIÓN FORZADA	2	0	0	2
19	DELITOS CONTRA LA LIBERTAD E INTEGRIDAD SEXUAL	2	0	0	2
50	ATENTADO	2	0	0	2
9	HOSTIGAMIENTO	2	1	0	3
46	VIOLACIÓN A LOS DDHH E INFRACCIÓN AL DIH	3	0	0	3
20	DESAPARICIÓN FORZADA	3	0	0	3

Vista preliminar de una base de datos en formato Excel

i Existen bases de datos muy pesadas. Por ejemplo, el Registro Único de Víctimas cuenta con aproximadamente 9 millones de registros, y eso dificulta abrir el archivo; para solucionar la limitación, el procedimiento que se sigue es dividir la información entre las diferentes hojas de Excel, con el fin de que cualquier persona la pueda consultar y utilizar el recurso desde este formato; de lo contrario, son bases de datos que deben ser descargadas en formatos TXT o CSV. Para el caso específico de la Comisión de la Verdad, las bases de datos que cuentan este proceso son: RUV, SPOA y Sistema de Información de Justicia y Paz de la Fiscalía.

❗ La visualización de los datos en el portal de microdatos depende 100% de cuántos registros tenga el archivo. No es posible visualizar, dentro del portal, una base de datos con un número significativo de registros.

- Los requerimientos de máquina (infraestructura), dependen de la cantidad de recursos almacenados a través del aplicativo. Al existir bases de datos de grandes tamaños, crece el consumo de recursos; esto puede llevar a que los servidores colapsen. A medida que aumenta el sistema, deben aumentar los recursos.
- El portal de microdatos permite visualizar tablas, mapas y gráficos; sin embargo, los mapas y los gráficos presentan ciertas limitantes para estructurar la visualización, ya que muchas de las bases de datos contenidas en el repositorio, no cuentan con una georreferenciación adecuada (solo cuentan con el nombre de un municipio, por ejemplo) para generar el mapa; y, muchas otras, al ser tan grandes, poseen demasiadas variables que hacen que su graficación sea compleja. En este sentido, estas dos posibilidades visuales no son aprovechadas por la herramienta.

i Lo que se realiza para suplir estas limitantes es definir unas tablas dinámicas que permiten interpretar la base de datos de manera más sencilla; no obstante, este ejercicio depende del proceso de investigación y de la base de datos. En otras palabras, es un proceso que se hace solo cuando es solicitado desde los equipos de investigación, ya que implica alterar las tablas originales (se quitan datos o se hace un proceso de depuración); en este sentido, dentro de los dataset no se encuentran tablas dinámicas, ya que es una alternativa para el ejercicio de investigación y no un recurso del dataset.


Las bases de datos que tienen el campo geoespacial, se disponen en el geoportal de la Comisión de la Verdad ([ver documentación de geoportal](#)).


- Cada dataset entregado a la Comisión de la Verdad se documenta a partir de los oficios y/o expedientes que acompañan su proceso de entrega y recepción. En la Comisión de la Verdad, estos documentos se archivan en el FileServer, clasificándolos de acuerdo con la entidad que entrega: entidad pública, entidad privada o Comisión de la Verdad (para las bases de datos originadas por la institución).
- Los dataset se clasifican, entre muchas otras cosas, a partir de licencias. Estas licencias establecen la manera en que cada dataset debe ser usado; es

decir, son las condiciones en las que el contenido del recurso puede ser utilizado, distribuido o reproducido. Para mayor información visitar: <https://opendefinition.org/licenses/>.

Recomendaciones para la interacción con el código fuente CKAN

- Es importante entender la arquitectura de CKAN, basada en la inyección de dependencias, con el fin de poder hacer modificaciones al código. Una de las limitantes tiene que ver con la complejidad de la estructura y la poca documentación existente al respecto. En efecto, la inyección de dependencias implica que para cambiar una funcionalidad, se debe crear una nueva que modifique la existente (no es posible modificar la funcionalidad directamente); esto conlleva a que el desarrollo se de en función de herencias que complejizan la estructura del aplicativo: las funcionalidades no desaparecen, sino que se ocultan y se ejecutan las nuevas instrucciones.

 La documentación de CKAN es poca y sencilla, lo que dificulta articularse con el código. Según el equipo de ingenieros, la documentación no habla de la estructura del código, sino de algunas capas: el núcleo donde se hacen las consultas, la capa donde se procesan los datos y finalmente la interfaz donde se visualizan.

 Por lo general las herramientas de software libre no entregan una gran documentación del código. Esto implica un trabajo adicional para el ingeniero desarrollador, el cual tiene que hacer un estudio de ese código que puede demandarle una importante inversión de tiempo.

- La versión de CKAN que se usó para el desarrollo del portal de microdatos de la Comisión de la Verdad, combina dos Framework para el desarrollo web: Flask y WZ Werkzeug. En este sentido, una de las dificultades fue estabilizar ambos Framework, en tanto una parte del desarrollo funciona en Flask y otra parte funciona en WZ Werkzeug. Adicionalmente, adaptar el plugin DDI tuvo también un nivel de dificultad en tanto era necesario adaptarlo para las especificidades de la Comisión de la Verdad.

Entornos y metodologías

Aplicativo

El portal de microdatos es una herramienta dispuesta en el software [CKAN](#), que es un software *Open Source* (código abierto); es decir, aunque el portal de microdatos no es un desarrollo *in house* como sucede con otras herramientas de la Comisión de la Verdad, tiene la ventaja de no generar costos de licenciamiento, y contar con un desarrollo colaborativo que permite implementar nuevas funcionalidades a partir del desarrollo de plugins.

El CKAN es una aplicación licenciada bajo la licencia GNU AGPL ([Aferro General Public License](#)), de autoría de [Open Knowledge Foundation](#).

Tecnologías y arquitectura

- CKAN; versión 2.8.4

<https://docs.ckan.org/en/latest/contents.html>

- Python; versión 3.8

<https://www.python.org/doc/>

- PostgreSQL; versión 9.5.19

<https://www.postgresql.org/docs/9.5/release-9-5-19.html>

- Indexador Solr; versión 4.7.2

<https://solr.apache.org/guide/>

Para el desarrollo web:

- Framework Flask; versión 1.1

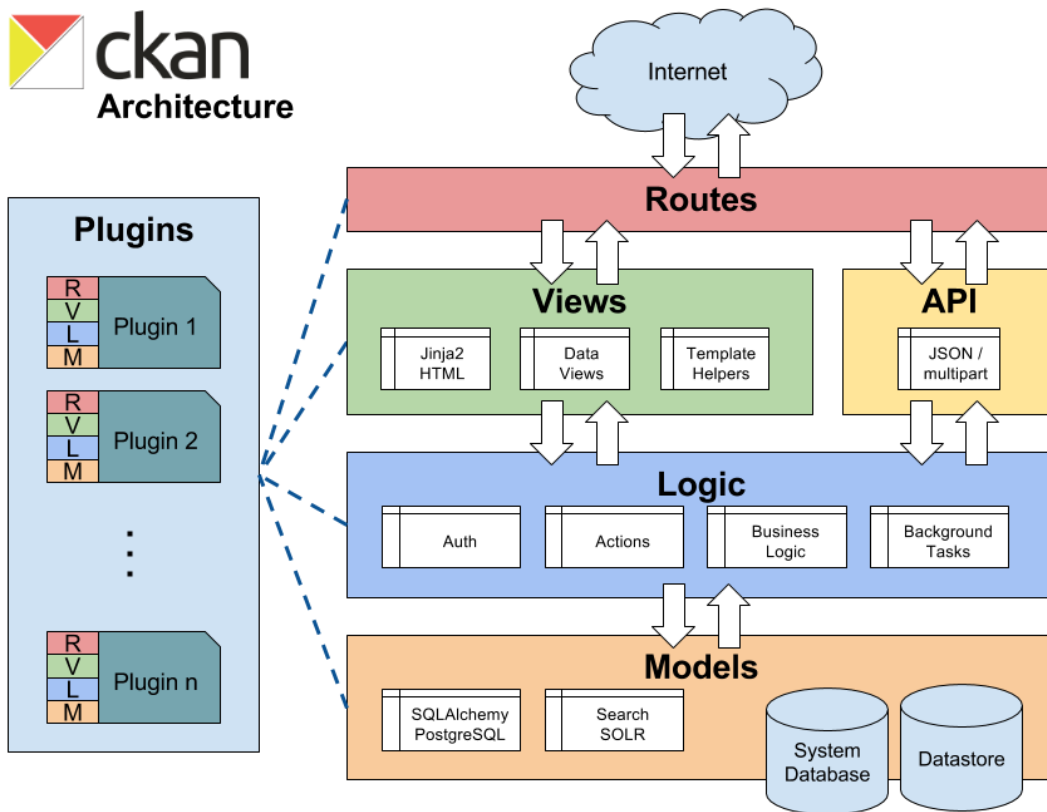
<https://flask-es.readthedocs.io/>

- Framework WZ Werkzeug; versión 1.0

<https://werkzeug.palletsprojects.com/en/2.1.x/>

- El proceso de anonimización de las bases de datos se hace a través de scripts de R y Stata.

- Los desarrolladores de la Comisión de la Verdad, hicieron modificaciones al código de Ckan e implementaron algunos plugins para mejorar y extender la aplicación. Entre los plugins implementados se encuentran:
 - Plugin DDI adaptado a la Comisión de la Verdad (con el fin de incorporar algunos campos de la [Data Documentation Initiative](#) -metadatos DDI-).
 - Plugin para hacer login con el directorio activo (plugin Ldap).
 - Plugin para dar permisos especiales de acceso a los dataset (para aquellos usuarios que solicitan acceder a bases de datos privadas no anonimizadas o con algún tipo de reserva) - plugin Extrapermissions.
- La modificación del código se hizo en función de:
 - Mejorar el rendimiento.
 - Organizar requerimientos específicos de la Comisión: licencias, niveles de acceso, público/privado, obligatoriedad o no de algunos campos, entre otros aspectos. No existe un plugin para estos ajustes, por tanto, se optó por modificar directamente el código, el núcleo donde ya venían esquematizados.
- Las modificaciones a CKAN en función de la Comisión de la Verdad se hicieron, en su gran mayoría, en el componente Logic, donde se encuentra toda la lógica de programación. En el componente Models se encuentran todas las conexiones a las bases de datos, allí se modificaron algunos campos con el fin de que fueran opcionales. Finalmente, dentro del componente Views se modificó la forma de presentar los datos.



Arquitectura CKAN

El componente Logic del portal de microdatos de la Comisión de la Verdad, se encuentra en:

`comisiondelaverdad/sim-ckan/-/tree/main/ckan/logic`

Copied!

Metodologías

- **Programación Sprint Scrum:**
 - Historias de usuario
 - Definición de tableros y actividades (sprints)
 - Planeación y seguimientos

Esta metodología se lleva a cabo a través de tableros Jira, donde se hace seguimiento a los sprints.

Buenas prácticas

Programación Scrum: se recomienda trabajar bajo la metodología de programación Sprint Scrum, dado que es una metodología que permite la articulación fácil y rápida de modelos simples y funcionales, los cuales pueden ir evolucionando con el feedback y las necesidades del cliente.

Bajo esta metodología estaremos en función de un desarrollo iterativo o incremental, el cual consiste en dividir el trabajo en pequeñas partes o bloques temporales (sprint). Al final de cada etapa se entrega una funcionalidad completa. Para estructurar la evolución se recomienda crear el Mínimo Producto Viable (Minimum Viable Product, MVP): producto con suficientes características para satisfacer a los clientes y proporcionar retroalimentación para el desarrollo futuro.

MVP: herramienta que permite aprender mientras se desarrolla. Gracias a la iteración, el producto evoluciona y se reduce el tiempo para la validación de nuevas ideas.

Estandarizar código: definir unas reglas de trabajo; es decir, estandarizar la manera en que se van a crear y llamar las funciones, los métodos, las variables, los atributos, etc. La normalización del código es fundamental para el mantenimiento óptimo del desarrollo.

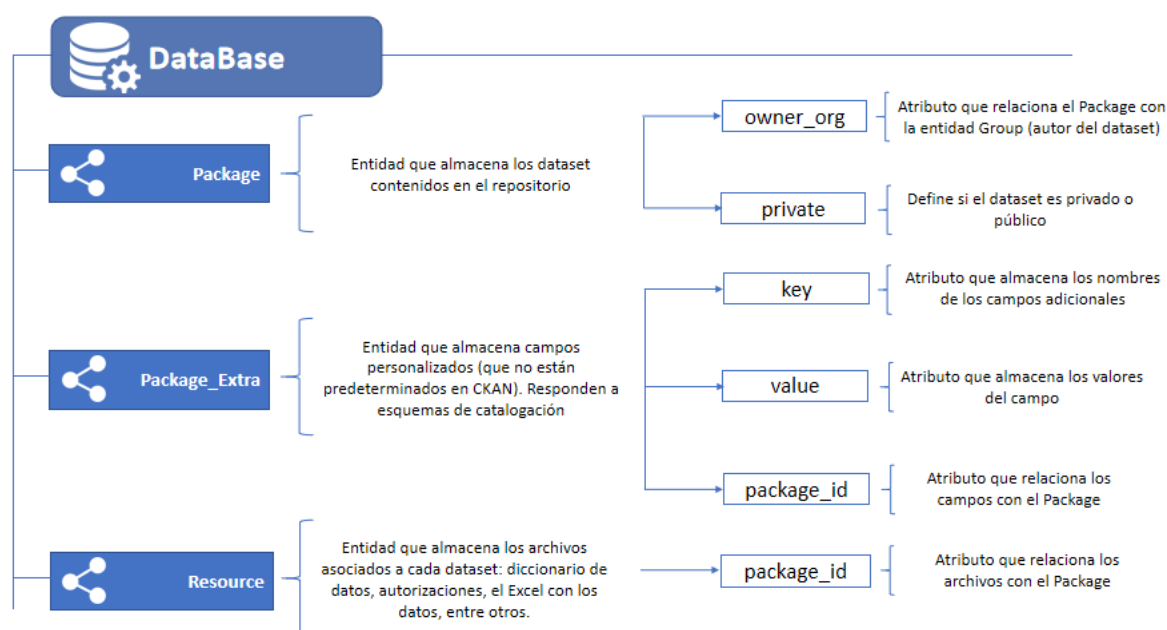
Comentar el código: es una buena práctica comentar el código a fin de que se convierta en un texto legible, autoexplicativo y facilite las modificaciones y el mantenimiento.

Interacción con usuarios: es una buena práctica tener entrevistas previas con los usuarios del sistema; dichos encuentros permiten entender los requerimientos puntuales y las funcionalidades concretas sobre las que debe basarse el desarrollo. Además, se recomienda encontrar un lenguaje común entre lo técnico del sistema y el requerimiento del usuario, a fin de no generar reprocesos y lograr comunicar fácilmente las funcionalidades del sistema.

Entidades DB

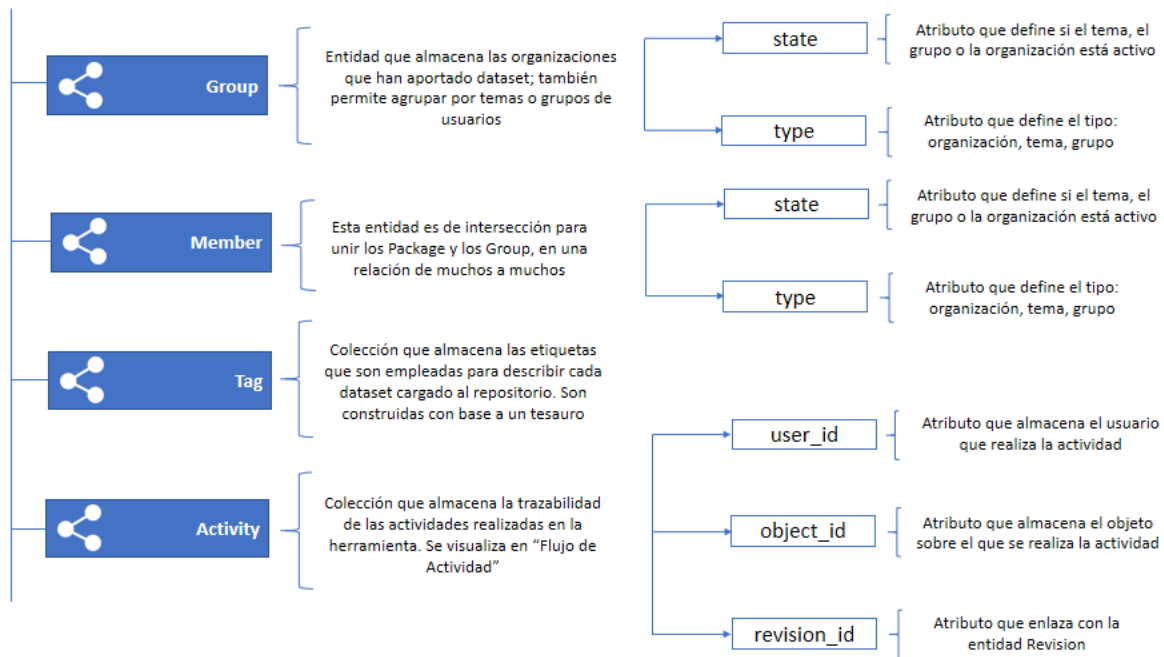
Arquitectura DataBase

Entidades usadas en el PostgreSQL



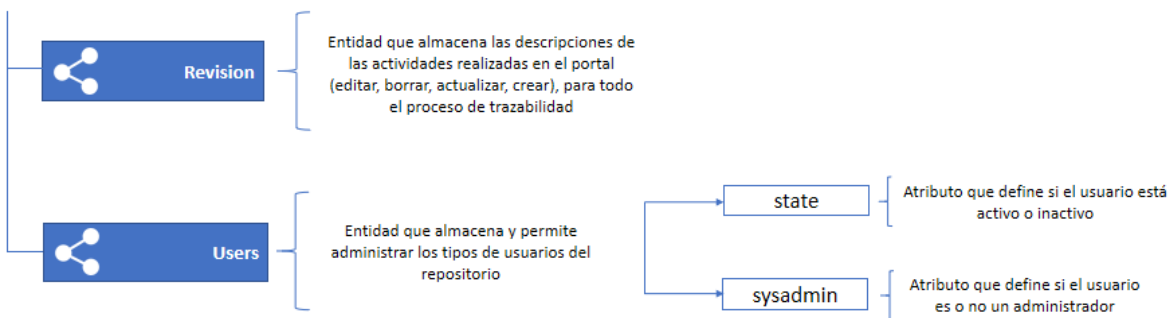
Arquitectura DB parte I

El PostgreSQL del portal de microdatos (CKAN) tiene como entidades principales el Package, el Package_Extra y el Resource; son entidades que se encargan de almacenar y ordenar los dataset que son aportados por diferentes entidades a la Comisión de la Verdad.



Arquitectura DB parte II

El siguiente grupo de entidades coadyuvan a la administración y disposición de los dataset; en este sentido, la entidad Group almacena las organizaciones, los temas y los grupos de usuarios con permisos especiales, y la entidad Member permite realizar la intersección entre ese Group y los dataset almacenados en el Package.



Arquitectura DB parte III

Desarrolladores

¿Por qué se ha logrado una buena herramienta?

De acuerdo con el equipo de ingenieros, el portal de microdatos es una buena herramienta porque permite centralizar, en un solo repositorio, un importante número de dataset pertenecientes a diversas entidades privadas y públicas del país. Dentro del Informe Final se referencian varias de estas bases de datos, ya que son una de las fuentes que acompaña los análisis y las conclusiones de los investigadores; en este sentido, la herramienta se convierte en un instrumento de contraste de lo establecido en el documento. En últimas, la Comisión realizó toda una labor de búsqueda de información, articulación, convenios interinstitucionales, con el fin de acceder a información clasificada que pudiera contrastar las hipótesis de investigación.

En ese sentido, si bien existe en el país un sin número de portales abiertos, no hay un repositorio que centralice el universo de datos existente, ya que cada entidad publica solo su información. La Comisión de la Verdad, en su portal de microdatos, ha centralizado la información de alrededor de 120 organizaciones públicas y privadas, lo cual convierte a la herramienta en el mayor portal de datos abiertos del país en función del conflicto armado:

Es un trabajo bastante grande; y si necesitan un portal de datos abiertos de información estructurada sobre conflicto armado en el país, creería que esta es la primer versión de lo que habría. Ni siquiera la Unidad de Búsqueda, ni la JEP tienen algo parecido frente a información estructurada; entonces, creo que eso sería un plus para nosotros (...) Es un buen inicio que seamos la única entidad del sistema que tiene un catálogo de microdatos, más allá de que nuestra duración sea más corta que la JEP y la Unidad de Búsqueda... Tener algo avanzado como CKAN, hace ver que la Comisión sí le paró bastante atención a todo lo de información estructurada que va a estar en el Informe...

Es este trabajo inmenso lo que llevó a replantear la versión de CKAN Andino implementada al principio. Según el equipo, fueron varios los inconvenientes, entre ellos el tema de seguridad, que obligó a readaptar otra versión de este software libre. Siguiendo lo planteado por los ingenieros, en el CKAN Andino todos los dataset quedaban públicos, incluyendo los que no estaban anonimizados (con información o datos personales); eso fue un motivo de alerta que obligó a realizar todo un proceso de migración y readaptación de la herramienta.

Aprendizajes y experiencias

Enfrentarse a un código abierto como el de CKAN conllevó una serie de experiencias y aprendizajes importantes para el conjunto de ingenieros desarrolladores. Para poder desarrollar el CKAN en función de las demandas de la Comisión de la Verdad, fue necesario un proceso de adaptación y acoplamiento con el núcleo del código; proceso que implicó varias horas de estudio y trabajo, y más si se tiene en cuenta que la documentación del CKAN es limitada.

Fue difícil al principio, me tocó adaptarme y entender el código de CKAN; poderme adaptar a ese código de CKAN para poder hacer las cosas... una vez pude entender cómo funcionaba CKAN, ahí sí pude realizar las cosas, los cambios que se iban a hacer...

De acuerdo con esto, y en términos del funcionamiento del portal, varios de los aprendizajes vienen dados en función del conocimiento y manejo de microdatos. En otras palabras, los ingenieros pueden estar familiarizados con los lenguajes de programación y las lógicas de desarrollo; no obstante, ello no implica que haya un conocimiento en el manejo de datos estructurados y/o microdatos, de allí que este aspecto se halla convertido, para varios de ellos, en un aprendizaje importante y necesario para llevar a buen término el desarrollo de la herramienta.

Ahora bien, trabajar con una herramienta como CKAN, que resguarda datos estructurados, implica tener contacto con otros equipos con el fin de definir aspectos como los metadatos, el etiquetado, entre otros aspectos; de hecho, uno de los desarrollos propios que se hicieron sobre el código de CKAN fue la reformulación de ciertos campos y la implementación del plugin DDI para los metadatos. De acuerdo con el equipo de desarrollo, se trabajó de la mano el equipo de lenguaje controlado para formalizar todo el tema de las etiquetas, ya que cada etiqueta asignada a un dataset debe estar previamente normalizada.

Asimismo, se realizó un trabajo colaborativo con el equipo de catalogación, para coordinar el tema de las licencias y los metadatos (estos últimos en función de la estructura del [metabuscador](#) y todo el tema de los campos DDI):

Debíamos establecer unas equivalencias entre CKAN y la estructura que se creó para el buscador. Entonces, allí hay unos esquemas que se definieron alrededor de los cuales hemos tenido que hacer ajustes en CKAN... Adicionalmente, no teníamos muy claro cómo iban a ser los campos DDI, había que buscar las equivalencias con personas que fueran expertas en el tema. Allí hubo unas diez reuniones para poder definir las y estar hablando el mismo lenguaje...

- ❗ Implementar un repositorio de bases de datos implica establecer parámetros sobre permisos de acceso, etiquetado y catalogación, licencias para el uso y reproducción de los datos, entre otros aspectos que deben ser tenidos en cuenta dentro del desarrollo de la herramienta. Este proceso conlleva un trabajo colaborativo con profesionales de diferentes áreas: lingüistas, catalogadores, analistas/estadistas, etc.

Expectativas

La expectativa más grande que existe dentro del equipo, es la posibilidad de que la ciudadanía se apropie de la información publicada, y sea el inicio de un trabajo importante alrededor de la verdad, más allá del Informe Final entregado por la Comisión.

- ✅ El proceso de anonimización llevado a cabo por la Comisión de la Verdad, se instaura como uno de los procesos más importantes dentro del portal de microdatos, en tanto es el proceso que permite que un mayor número de bases de datos sean públicas.

A nivel del código, se dejan publicadas las modificaciones realizadas a CKAN, con el fin de que se pueda descargar y actualizar la herramienta; adicionalmente, se espera que pueda seguir mejorando en términos de rendimiento y/o tipos de catalogación.

Usuarios

El equipo de analistas de datos de la Comisión de la Verdad, es el equipo que ha interactuado de mayor manera con el portal de microdatos, en función de preguntas e hipótesis perfiladas por la investigación; de acuerdo con el equipo, los datos dispuestos en el repositorio, ayudan a verificar o rechazar hipótesis, formular nuevas preguntas y darle soporte estadístico y cuantitativo a ciertas afirmaciones. Es un trabajo que no sería posible si no se pudiera acceder a las bases de datos de organizaciones públicas y privadas del país.

Experiencias: aprendizajes, retos y oportunidades

Las estimaciones de violencia fueron uno de los principales temas trabajados desde la analítica de datos en la Comisión de la Verdad; el homicidio, la desaparición forzada, el secuestro, el reclutamiento y el desplazamiento forzado, se perfilan como las cinco violencias predominantes dentro del proyecto de estimaciones de violencia. De acuerdo con el equipo, para esta labor fueron usadas alrededor de 120 bases de datos tanto de organizaciones públicas como privadas. Ahora bien, adicional a las estimaciones de violencia, el equipo también usó el repositorio para dos retos de investigación importantes: primero, entender la relación entre las dinámicas hidroeléctricas, agroindustriales y mineroenergéticas con el conflicto armado colombiano, en función de perfilar otras dinámicas económicas no necesariamente asociadas con los cultivos de uso ilícito; y segundo, la estimación del asesinato de líderes sociales desde 2002 hasta 2019, con el fin de entender el aumento de esta clase particular de homicidios después de la firma de los Acuerdos de Paz con las FARC.

De acuerdo con este proceso, el flujo se dio en función de preguntas e hipótesis de investigación formuladas por el equipo de investigadores de la Comisión de la Verdad; es decir, los investigadores tienen preguntas que son trasladadas a los analistas de datos que, a partir del ejercicios estadísticos, corroboran o rechazan las hipótesis:

Uno trabaja a partir de hipótesis. Entonces, por ejemplo, el caso de los líderes sociales... los investigadores tienen una idea: con el Acuerdo de Paz que se firmó con las FARC en el 2016, los líderes sociales son más victimizados, hay más asesinatos, más persecución contra el movimiento social. A uno le dicen esa hipótesis, y lo que uno tiene que corroborar es, a partir de las fuentes de información que tenemos, que eso es verídico... Entonces, de acuerdo con esa hipótesis, la

conclusión fue que, efectivamente, después del 2016 el asesinato de líderes sociales aumentó...

Ahora bien, este proceso de análisis de datos en función de hipótesis y preguntas, fue perfilando diferentes estrategias; una de ellas, por ejemplo, es la necesidad de solicitar a varias entidades y organizaciones sus bases de datos para engrosar el repositorio y tener un mayor número de fuentes. De esta manera, el equipo de analistas de datos estructuró todo un proceso de documentación para la solicitud y el resguardo de estas fuentes de información.

Con respecto al manejo del portal de microdatos, existe un conceso en función de la facilidad de la herramienta para ser usada y para cargar las bases de datos, a diferencia de herramientas como el [módulo de catalogación colaborativa](#), que sí requiere un mayor aprendizaje para cargar los documentos.

Logros y expectativas

Uno de los logros más importantes señalados por el equipo, es la naturaleza misma del portal de microdatos: centralizar en un único repositorio los datos de un número importante de organizaciones sociales e instituciones públicas y privadas; este hecho permite que las víctimas, a partir de sus propios registros, presenten sus versiones de lo que ha pasado y de lo que han vivido.

A mi me parece fundamental, dentro de este legado que nos deja el repositorio de CKAN, que tenemos múltiples fuentes de organizaciones sociales, de instituciones a las que también ha sido difícil el acceso... muchas veces se cuestionaron para que entregaran información a la Comisión de la Verdad. El tener esa información dispuesta en un solo repositorio, y tener acceso para investigar y trabajar, ha sido fundamental...

Este hecho, según el equipo, deja logros a futuro: la posibilidad de aproximarse a los datos para construir política y reparación hacia las víctimas. Ahora bien, es importante resaltar dos de los hallazgos más importantes logrados a partir de las bases de datos resguardadas en el repositorio: el aumento de asesinatos de líderes sociales después de la firma de los Acuerdo de Paz con las FARC; y la relación entre la explotación minera y el conflicto armado. Con respecto a este último punto, resalta el equipo, se relacionaron los indicadores de violencia con los indicadores de desplazamiento forzado teniendo como factor de causalidad la concesión de títulos mineros hace aproximadamente 12 años en el norte del Tolima; es decir, la hipótesis planteada por los investigadores se centró en afirmar que los títulos mineros que están en funcionamiento en este momento en el norte del Tolima, fueron despojados hace 12 años y se correlacionan con los indicadores de violencia y desplazamiento forzado en esa zona durante ese rezago temporal:

Aquí trabajamos varios datos, uno de ellos fue el de desplazamiento forzado, otros indicadores de violencia también dispuestos en el repositorio en el que hacíamos ejercicios de autocorrelación espacial para ver municipios específicos en donde hay una relación entre alta minería y alta violencia, y, efectivamente, se veía una ruta por el Pacífico... la ruta desde Nariño, por el Bajo Cauca antioqueño, por el sur de Bolívar, señalaba esos municipios donde había una dinámica marcada dentro de la violencia (entendida como indicador de violencia), y la violencia entendida como agregado de hechos victimizantes; y eso nos develaba mucha relación entre esa fuente que puede existir dentro de lo mineroenergético con el conflicto armado...

En este sentido, y siguiendo lo señalado por el equipo, las dinámicas económicas del conflicto, más allá de las economías ilegales, dan pautas para entender que el conflicto armado tiene expresiones que se usan en la legalidad y que son apoyadas por los gobiernos. Asimismo, y como valor agregado, el portal de microdatos es una herramienta que permite construir preguntas que, de otra manera, serían difícil formularse; en efecto, gracias a las bases de datos construidas a partir de las sentencias de restitución de tierras, fue posible marcar preguntas alrededor de los cultivos transitorios y permanentes y su relación con el conflicto armado, en claro contraste con los casos de monocultivo, como la palma, que han sido documentados de manera más exhaustiva.

Retomando todo estos aspectos, concluye el equipo, el portal de microdatos es una herramienta que permite, a partir de pautas cuantitativas, realizar ejercicios de esclarecimiento poniendo en evidencia, con datos concretos y fuentes verídicas, cómo fueron -y cómo siguen siendo- las dinámicas del conflicto. Adicionalmente, es una herramienta que permite generar contrastes con datos cualitativos; en efecto, los ejercicios de escucha de la Comisión, donde se genera un panorama no cuantitativo de los impactos y afrontamientos de las víctimas, permiten construir matices: entender no solo el agregado de hechos victimizantes, sino cómo esas cifras han marcado a las víctimas en sus relaciones cotidianas y en sus resistencias (cómo han afrontado el conflicto y cómo se han movilizado). Este aspecto hace del software una herramienta útil, en tanto permite ese contraste cuantitativo que toda investigación debe llevar.

Aportes a la ciudadanía

Según el equipo de analistas, el aporte más importante de la herramienta a la ciudadanía es la potencialidad que tiene no solo para entender el conflicto armado, sino para hacer ejercicios de política relevantes y construir dinámicas para la reparación de las víctimas. La información permite, por ejemplo, identificar los municipios donde se ha sentido con mayor fuerza el conflicto, para así generar una lectura más integral de país que, de otra manera, sería más complejo. En consecuencia, se pueden generar políticas sociales y de reparación focalizadas y

diferenciadas en función de la territorialidad del conflicto. Ahora bien, con respecto a la territorialidad del conflicto, anota el equipo, la herramienta permitió hacer zoom a varios factores, siendo el tema del territorio uno de los más importantes para entender el inicio y la permanencia del conflicto; en este sentido, y en concordancia con lo que plantean los analistas, probablemente una reforma integral agraria pueda ser el inicio de un posible fin del conflicto:

Yo creo que trabajando por la repartición de la tierra, o empezar a enfocarnos por tener un catastro multipropósito que nos de la información que tienen los municipios, los baldíos, que tenga información clara sobre quien tiene la tierra... nos llevaría a una dinámica para acabar con este conflicto. En lo fundamental, si no hay un ejercicio de reforma agraria o de lo que se cumpla con lo firmado en el Acuerdo de Paz en el punto de Reforma Rural Integral, yo creo que la guerra va a seguir existiendo...

Finalmente, anotan los analistas, el valor agregado de esta herramienta viene dado en el sentido de que el portal de microdatos de la Comisión de la Verdad, es el único repositorio que centraliza un número importante de bases de datos de diferentes organizaciones e instituciones del país (públicas y privadas) que caracterizan el conflicto armado y las víctimas:

Yo creo que este repositorio que tenemos tiene mucho potencial por eso: al ser las mismas organizaciones sociales las que han presentado informes y sus bases de datos de reconocimiento de sus víctimas da esa importancia; escuchar a las organizaciones sociales escuchando a las víctimas, por eso me parece importante... Cosa que no hace la Fiscalía, la Fiscalía indaga lo que quiere; y esto tiene una dinámica más de ver con la organización y su forma de leer cómo pasaron las cosas...

Posibilidades de mejora

- El equipo apunta que la herramienta puede mejorar en términos de vistas; es decir, poder generar visualizaciones sobre los datos para dar una mirada global de la información. Durante el proceso en la Comisión de la Verdad, esta demanda fue suplida con lo que ofrecían los [tableros de visualización](#)