



Anteproyecto

# Plan Nacional de Búsqueda

Hito 3



16/11/2023

<b>1</b>	<b>Plan de trabajo.....</b>	<b>6</b>
1.1	Septiembre: Hito 1 .....	6
1.2	Octubre: Hito 2.....	7
1.3	Noviembre: Hitos 3 y 4 .....	8
1.4	Stakeholders y sus procesos actuales .....	9
1.5	Sistemas utilizados actualmente .....	13
<b>2</b>	<b>Arquitectura actual .....</b>	<b>16</b>
<b>3</b>	<b>Mapa de calidad de datos.....</b>	<b>17</b>
3.1	Metodología .....	17
3.2	Diagrama del proceso .....	18
3.3	Definición de calidad .....	19
3.4	Ejemplo de segmentos en una página de un documento.....	19
3.5	Estadísticas para 100 documentos .....	22
3.6	Tiempo de procesamiento y costo.....	24
3.7	Ejemplos de páginas con mala calidad de transcripción.....	25
3.8	Próximos pasos .....	27
<b>4</b>	<b>Propuesta de módulos, sistemas y componentes a construir .....</b>	<b>28</b>
4.1	Módulo de documentos.....	28
4.2	Módulo de procesamiento.....	28
4.3	Módulo de investigación .....	29
4.4	Módulo de anotaciones .....	30
4.5	Módulo de trayectorias.....	30
4.6	Módulo de familiares .....	30
4.7	Terreno.....	31
4.8	Jurídica.....	31
4.9	Servicio Médico Legal.....	31

<b>5</b>	<b>Arquitectura final.....</b>	<b>32</b>
5.1	Principales Características .....	33
5.2	Componentes .....	34
<b>6</b>	<b>Definición base de datos.....</b>	<b>36</b>
6.1	Modelo de Usuarios.....	36
6.2	Modelo Repositorio.....	37
6.3	Modelo cola de procesamiento.....	38
6.4	Modelo eventos de dominio .....	38
6.5	Modelo de objetos.....	39
6.6	Modelo de grafo.....	40
<b>7</b>	<b>Interoperabilidad .....</b>	<b>42</b>
7.1	Sistemas Internos.....	42
7.2	Sistemas Externos .....	42
<b>8</b>	<b>Wireframes y flujo de navegación .....</b>	<b>44</b>
<b>9</b>	<b>Definición de usuarios.....</b>	<b>44</b>
<b>10</b>	<b>Maquetas .....</b>	<b>45</b>

# 1 Plan de trabajo

A continuación, se presenta la hoja de ruta para el anteproyecto del Plan Nacional de Búsqueda separada por cada mes de trabajo y con la descripción de cada hito.

## 1.1 Septiembre: Hito 1

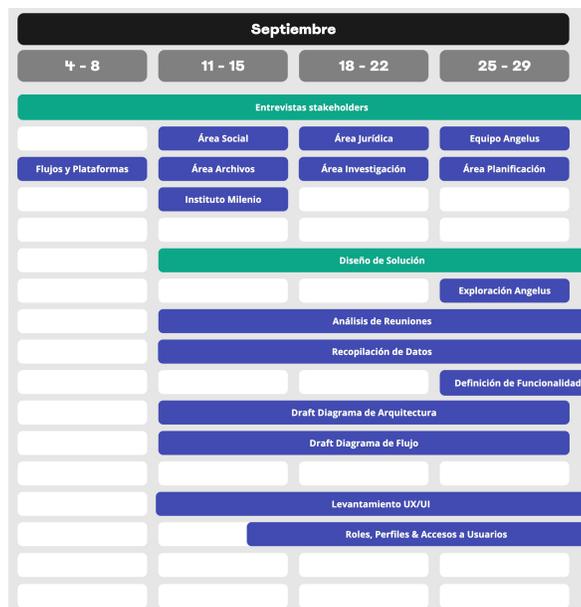


Ilustración 1. Hoja de ruta mes de septiembre.

Durante el primer mes de trabajo la actividad central será la celebración de reuniones con las distintas áreas involucradas en el Plan Nacional de Búsqueda (Stakeholders). En paralelo el equipo de Unholster comenzará a diseñar la solución.

El entregable es el plan de trabajo, requisito para dar cumplimiento al hito 1 del proyecto.

## 1.2 Octubre: Hito 2

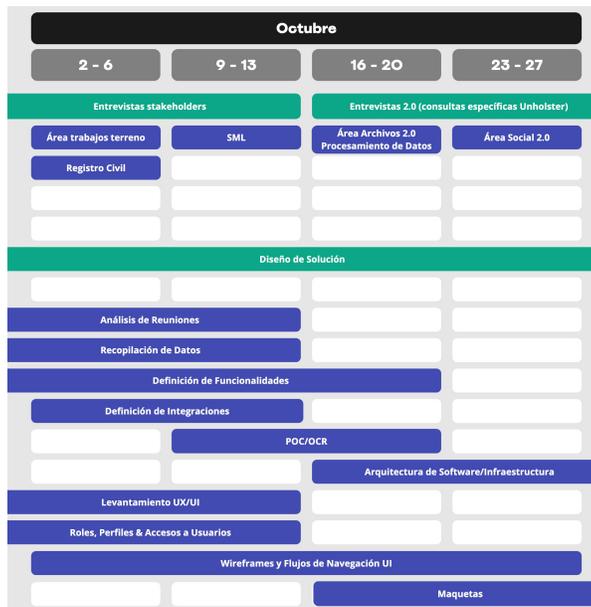


Ilustración 2. Hoja de ruta mes de octubre.

Durante la primera mitad de octubre se sostendrán las últimas reuniones con los *Stakeholders* para cumplir la tercera semana con el hito 2 que considera entregar:

- Esquematización de procesos actuales
- Arquitectura actual
- Mapa de calidad de los datos
- Interoperabilidad - identificación sistemas internos y externos existentes que han de integrarse
- Arquitectura final V0

### 1.3 Noviembre: Hitos 3 y 4

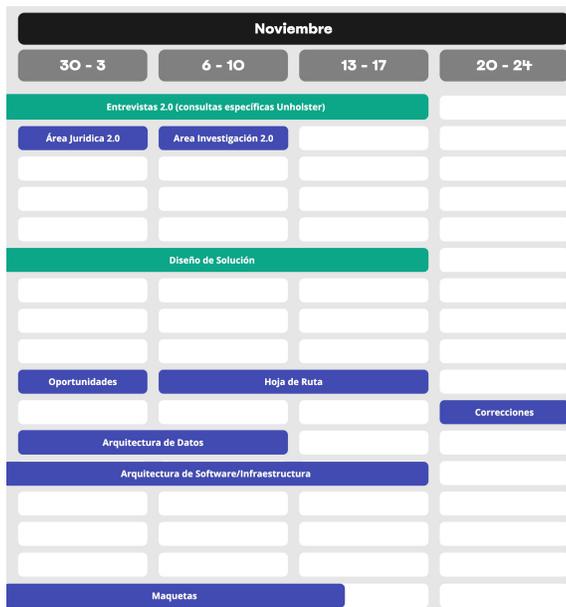


Ilustración 3. Hoja de ruta mes de noviembre.

A lo largo del último mes se trabaja en los siguientes puntos, cumpliendo el hito 3 para la segunda semana:

- Propuesta de módulos, sistema y componentes a construir
- Arquitectura final

Para los módulos de Documentos y Contenidos:

- Definición de base de datos
- Wireframe y flujo de navegación
- Definiciones de usuarios
- Maquetas

Finalmente, para la cuarta semana se trabaja en los entregables finales considerando las correcciones acordadas con el cliente:

- Definición del MVP
- Plan de implementación de mejoras
- Oportunidades
- Esquematización de procesos actuales

## 1.4 Stakeholders y sus procesos actuales

A continuación, se presentan un conjunto de *stakeholders* o áreas que tendrán injerencia en la Plataforma Nacional de Búsqueda con las cuales se sostuvo al menos una reunión durante el primer mes del anteproyecto.

### 1.4.1 Área archivos

Son la principal fuente de información histórica cuyo rol es pasar a formato digital una gran cantidad de documentos físicos albergados en el ministerio. Trabajan asistidos por un sistema llamado Alexandria (<https://www.goalexandria.com/about/>), el cual permite caracterizar documentos y conocer su ubicación en los servidores donde finalmente se almacenan. Además, son usuarios del sistema de derechos humanos.

Las principales necesidades levantadas por esta área corresponden a:

- Gestión Documental:
  - Gestión de digitalización de documentos.
  - Identificar con seguridad y rapidez el origen físico de la digitalización de un documento.
  - Poseer metadatos<sup>1</sup> de los documentos digitalizados.
  - Migrar desde la norma bibliotecológica MARC21 a ISAD(G).
  - Considerar la configuración de ISAD(G) como patrón de metadatos en el gestor documental.
  - Considerar la configuración de MARC21 como patrón de meta datos en el gestor documental para soportar los datos bibliotecológicos durante el proceso de migración a ISAD(G).
- Proceso de curación de la digitalización.
- Aplicación de inteligencia artificial para poder etiquetar textos, obtener resúmenes y procesar digitalizaciones.
- Poder realizar búsquedas sobre los documentos.
- Presentar cierto contenido al público general.
- Repositorio digital unificado.
- Flujo automatizado de digitalización para documentos.

Los desafíos que presenta el área son:

- Entendimiento del volumen actual de información que poseen.

<sup>1</sup> Entendemos por metadatos como cualquier información que no pertenece al documento pero si lo describe y/o caracteriza. Son datos sobre los datos que contiene el documento.

**Comentado [CH1]:** La metadata es solo un ámbito de la información que se requiere recuperar en el marco de la descripción documental, entre otros. Es importante recuperarlo pero con ello no se cumple a cabalidad las tareas referidas a este asunto. Importante entender que el software Alfresco recupera la metadata usando una norma llamada Dublincore que no es la que se utilizará. La que usaremos es ISAD (G).

**Comentado [CH2]:** Considerar la recuperación de lo catalogado en formato Marc21 que será traducido a la norma ISAD (G) de descripción archivística. En cuanto a la digitalización, recuperación de información y eventual automatización del poblamiento de los campos en ISAD (G), esto debe suponer un proceso de validación ex post.

**Eliminó:** el formato MARC21 y ISAD(G) para datos bibliográficos

**Eliminó:** <#>Migración de metadata en formato MARC21 a formato ISAD(G)

**Eliminó:** <#>¶

- Considerar la existencia de documentos tipo texto, imágenes, audios, mapas y videos.
- Calidad de los documentos a digitalizar.
- Análisis de las opciones de OCR del mercado.
- Definir criterio para la organización de la información existente.

**Comentado [CH4]:** + audios, mapas y otros.

**Comentado [CH5]:** Los acervos que se subirán no están organizados aún, es necesario considerar que se puedan ordenar ex post. ¿Están considerando algún criterio que permita organizar la documentación subida a la plataforma?

#### 1.4.2 Área social

Es la encargada de gestionar la relación, comunicación y coordinación con los familiares de las víctimas. Considerando el acompañamiento individual y colectivo, diligencias judiciales, mediaciones entre las familias, coordinación de trabajos en terreno, entre otros. Su principal soporte es el sistema DDHH donde cuentan con su propio módulo.

Las principales necesidades levantadas por esta área corresponden a:

- Disponibilidad de una base de datos de los familiares con información de contacto.
- Integración con el SRCel (esta funcionalidad no se puede garantizar en la medida que no depende de Unholster tener acceso a los servicios del SRCel).
- Seguimiento de los trámites y consultas digitales solicitados.
- Seguimiento de los acompañamientos y mediaciones realizadas con una bitácora.
- Gestión de los encuentros nacionales y mesas de trabajo.
- Gestión y planificación de los trabajos en terreno.
- Gestión de financiamiento para familiares.
- Comunicaciones automatizadas y *newsletter*.
- Robustecer las mallas familiares con información emergente (agregar datos de familiares nuevos)
- Contar como mínimo con la capacidad de búsqueda y cruce que actualmente se utiliza en Sistema de DDHH.

**Comentado [CH6]:** Incorporar posibilidad de interoperatividad en caso de ser preciso. Por ejemplo, en este punto con Registro Civil

Los desafíos que presenta el área son:

- Validar la calidad de información de contacto de los familiares existentes.
- Definir criterios para presentar información sensible a familiares.

#### 1.4.3 Área investigación

Es la consumidora de los datos existentes para generar hipótesis de trayectorias de las víctimas definidas en la Comisión Nacional de Verdad y Reconciliación (Rettig); la Corporación Nacional de Reparación y Reconciliación y la Comisión VALECH 2. Además de otras que puedan aparecer en el futuro que provean datos relevantes para el Plan Nacional de Búsqueda.

**Comentado [CH7]:** Además de otras que se puedan reconocer en el transcurso de su operación.

Las principales necesidades levantadas por esta área corresponden a:

- Poder generar múltiples hipótesis de trayectorias de víctimas, es decir, construir la historia desde que es detenido hasta, idealmente, la identificación de sus osamentas.
- Considerar como inicio de una hipótesis una víctima u osamentas encontradas, por tanto, poder reconstruir la historia de principio a fin.
- Dentro de una trayectoria, de existir, se debe poder identificar uno o más:
  - Personas: víctimas, testigos, agentes, osamentas.
  - Lugares: recintos, lugar de inhumación, sitios de interés.
  - Hechos: captura, **detención**, traslado, tortura, asesinato, entre otras.
  - Organizaciones: políticas, religiosas, de Inteligencia, Fuerzas Armadas, entre otras.
  - Temporalidad: fechas, duraciones relativas.
- Cada trayectoria tiene una estructura o ficha base que se desea reconstruir.
- Debe permitir abrir casos contra los agentes involucrados y hacer seguimiento de estos.
- **Gestión de hipótesis y su relación con patrones marcocriminales emergentes.**

Los desafíos que presenta el área son:

- Presentación de datos en formato de grafos, permitiendo la creación de estos en tiempo real.
- Revisión y búsqueda entre los documentos existentes en el sistema de manera eficiente y con ayuda de inteligencia artificial.
- Poder explorar datos no uno a uno si no en conjunto (revisar lugares, organizaciones, conjunto de víctimas, etc.), apoyado en gráficos, reportes y fichas.
- **Curar contenido de documentos agregando comentarios, dejando un registro trazable en el tiempo de los cambios y sus autores.**

#### 1.4.4 Área jurídica

Utilizan principalmente el Sistema de Derechos Humanos en conjunto con la información que se va subiendo al Poder Judicial ( <https://www.pjud.cl/> ) respecto al avance que van teniendo las causas judiciales abiertas contra los distintos agentes ingresando nueva información al sistema de DDHH.

Las principales necesidades levantadas por esta área corresponden a:

- Identificar una causa por cualquiera de sus roles, los cuales van cambiando conforme avanza el proceso.
- Conocer el último estado procesal de una causa y enlazar toda su información con los documentos existentes con lo cual se tiene trazabilidad de la causa.
- Poder realizar búsquedas y filtros sobre las causas.

Los desafíos que presenta el área son:

**Comentado [CH8]:** Es detención

**Eliminó:** retención

**Comentado [CH9]:** sumar: hechos, organizaciones y temporalidad.

**Comentado [CH10]:** El proceso de formulación de hipótesis se encuentra asociado a la emergencia de patrones marcocriminales, por lo tanto, la plataforma debe poder relacionar ambas cosas.

**Eliminó:** s.

**Comentado [CH11]:** La plataforma debe incorporar la posibilidad de codificar y agregar comentarios al contenido de los documentos, dando cuenta del usuario que realizó tal acción.

**Comentado [CH12]:** y a través de distintas alternativas de representación: gráfico, informe o ficha de víctima, etc.

**Con formato:** Párrafo de lista

**Eliminó:** ¶

- Tener una comunicación con el área social respecto al avances de las causas generando un sistema de alertas.

#### 1.4.5 Servicio de Registro Civil e Identificación

Este servicio público provee información básica para las investigaciones realizadas, ellos poseen datos de hechos vitales y relaciones familiares, los cuales tienen respaldo en documentos físicos como fuente de verdad. Actualmente la obtención de esta información es por solicitud por medio de su página web o correo, se espera poder hacer más eficiente esta comunicación con una integración directa a la API del SRCel.

Las principales necesidades de información corresponden a:

- Conocer los hechos vitales de personas (víctimas, testigos, agentes).
- Conocer relaciones familiares de personas (víctimas, testigos, agentes).
- Poder identificar a una persona sin RUT.
- Realizar búsquedas por nombre y RUT.

Los desafíos que presenta el SRCel son:

- Existen personas que no poseen RUT y no es factible legalmente asignarle uno.

#### 1.4.6 Servicio Médico Legal

Este servicio realiza trabajos como parte en diversas actividades y diligencias relacionadas a las víctimas del programa, generando nueva información, principalmente de avances en las investigaciones, la cual debe ser incorporada al sistema. Actualmente se mantienen conversaciones con el SML para conocer el estado de cada investigación, pero se busca que a futuro se pueda tener una integración directa con entre los sistemas del PNB y los sistemas del SRCel.

Las principales necesidades de información corresponden a:

- Estatus de víctimas.
- Resoluciones judiciales.
- Toma de muestras (donadores de ADN).
- Datos antropomórficos.
- Lugares de interés.
- Actividades periciales.
- Inventario osamentas.

Los desafíos que presenta el SML son:

- Asignación de una persona para el ingreso de información.
- Conversaciones legales de entrega y presentación de la información.

### 1.4.7 Angelus

Sistema de referencia de origen mexicano utilizado como referente para la digitalización de documentos, etiquetados descriptivos de los mismo y visualización en forma de grafo de las relaciones previamente construidas a través del etiquetado.

Resuelven varias de las interrogantes y necesidades del sistema para el Plan Nacional de Búsqueda. Es un referente para estimar el tiempo de desarrollo que toma generar una plataforma de esta naturaleza y complejidad, así como también sirve como base para construir el sistema de apoyo al Plan Nacional de Búsqueda, principalmente en temas de funcionalidades, visualizaciones y bases de datos.

## 1.5 Sistemas utilizados actualmente

Durante las conversaciones con las distintas áreas se identificaron los siguientes sistemas que se utilizan en la labor diaria de cada funcionario.

### 1.5.1 Sistema de derechos humanos

Se trata de un sistema de diseño interno con desarrollo subcontratado que permite la recopilación y consulta de información relacionada con todas las entidades involucradas en casos de violación de los Derechos Humanos, incluyendo a las víctimas, victimarios, lugares, juicios, entre otros.

Este sistema cuenta con una sólida estructura de categorías conforme al trabajo y criterios históricos del Programa de Derechos Humanos, que abarcan:

- Personas
- Relaciones
- Lugares
- Procesos
- Fechas

No obstante, la aplicación carece de un diccionario de datos y códigos que permita relacionar la información que reside en la base de datos.

En el proceso de desarrollo de la plataforma, se han identificado diversos desafíos que es importante tener en cuenta, los cuales se pueden ver materializados en el sistema de Derechos Humanos actual:

- Falta de un enfoque adecuado en cuanto a la auditabilidad de las acciones realizadas en la plataforma.

**Comentado [CH13]:** no es conforma al PNB sino al trabajo y criterios históricos del Programa de Derechos Humanos. El PNB es parte del Programa.

**Con formato:** Normal, Sin viñetas ni numeración

**Comentado [CH14]:** el sistema actual es la herramienta que tenemos y nos ayuda muchísimo, sin embargo, carece de un libro de códigos que permita relacionar adecuadamente las entidades de la bbdd.

**Eliminó:** <#>Limitada disponibilidad de documentos digitalizados en el sistema.¶

- Excesiva dependencia de procesos manuales.
- Carencia de documentación técnica que respalde el funcionamiento del sistema.
- Interfaz de usuario no intuitiva y deficientes aspectos de experiencia de usuario (UX/UI).

Sus principales funcionalidades corresponden a:

- Configuración y Administración General del Sistema:
- Operaciones:
  - Víctimas
  - Testigos
  - Agentes
  - Recintos
  - Posibles Víctimas
- Procesos Judiciales
- Obras de Reparación Simbólica
- Atención Social a Familiares
- Registro de documentos
- Informes

### 1.5.2 Alexandria

Alexandria se destaca como un software de automatización de bibliotecas multiplataforma utilizado alrededor del mundo. Sin embargo, presenta ciertos desafíos a considerar. Alexandria no es un repositorio, es un catálogo, por lo tanto, solo sirve para contener meta data sobre los textos y no los textos en sí mismo, dado que estos residen en un repositorio independiente. Este trabajo de catalogación es manual, lo que redundará en una carga de trabajo adicional para los administradores del material, con nula asistencia para la mitigación del riesgo de error. Además, el sistema se encuentra desacoplado de otros sistemas, lo que dificulta la integración con herramientas complementarias.

Entre sus propiedades soporta la descripción bibliotecológica MARC21, la que se usa para la gestión de material bibliográfico. Norma que debe ser reemplazada por ISAD(G) que será incorporada a los procesos archivísticos del Programa de Derechos Humanos, y por tanto, a todo el material que se recopile en el marco del Plan Nacional de Búsqueda, razón que justifica el reemplazo de este sistema.

### 1.5.3 Angelus

Pese a no utilizarse actualmente, es un referente para la plataforma a construir. Es un sistema computacional mexicano de almacenaje y sistematización de documentos que contiene información sobre víctimas de desaparición forzada presenta un enfoque integral en la gestión de estos

**Comentado [CH16]:** Si bien podemos cambiar esta descripción es una buena forma de decir que sirve para guardar referencias a los textos y no los textos en sí mismo.

**Eliminó:** cuenta con catalogación archivística en la norma MARC21

**Eliminó:** en el caso del

**Eliminó:** por la ISAD(G)

**Eliminó:** por la cual

**Eliminó:** se debe r

**Eliminó:** ar

**Comentado [CH18]:** No se trata de una herramienta para la descripción archivística. Por eso necesitamos trasladarnos a la ISAD (G)

**Comentado [CH19]:** no se ingresan documentos pues no es un repositorio. Estos sí pueden caracterizar en base a una serie de campos que son más flexibles que los del SDH.

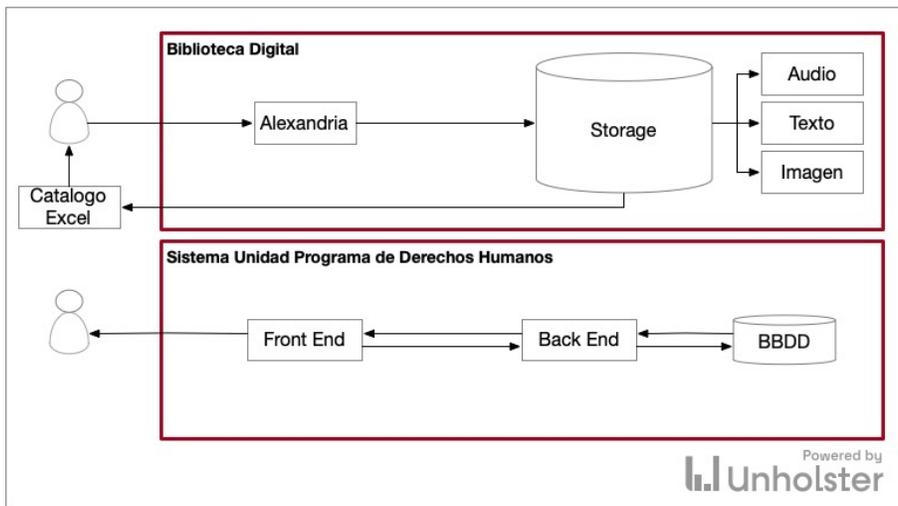
**Comentado [CH20]:** La norma será reemplazada e incorporada a los procesos archivísticos del Programa de DDHH y, por tanto, de todo lo que se recopile en el marco de Plan Nacional de Búsqueda.

datos. Uno de sus principales puntos a favor es la ontología, la cual se encuentra muy bien definida, facilitando la organización y recuperación de información de manera eficiente. Sin embargo, enfrenta ciertos desafíos que requieren atención. Por un lado, se encuentra la alta manualidad en los procesos, lo que puede ralentizar la digitalización y etiquetado de documentos, aumentando la posibilidad de errores humanos. Además, el sistema carece de un enfoque de auditabilidad, lo que dificulta el seguimiento y la verificación de las acciones realizadas en el sistema.

En cuanto a sus funcionalidades principales, el sistema permite el ingreso de documentos de manera efectiva, facilitando la incorporación de nueva información sobre las víctimas de desaparición forzada. También ofrece herramientas de anotación que permiten enriquecer la información presente en los documentos. Además, destaca por su visualizador de relaciones, que representa de forma gráfica las conexiones previamente construidas a través del etiquetado, lo que contribuye a una comprensión más completa de los datos y las relaciones entre las víctimas y otros elementos relevantes.

## 2 Arquitectura actual

Actualmente podemos identificar dos sistemas que no conversan entre sí, descritos en [Sistemas utilizados actualmente](#). El siguiente diagrama describe la situación actual:



### 3 Mapa de calidad de datos

En acuerdo mutuo se decide utilizar el disco duro compartido como base del análisis de los datos a considerar en el proyecto. Dicho disco duro se compartió el 21 de septiembre y fue devuelto el 12 de octubre.

El disco duro compartido tiene en total 95.707 archivos, de los cuales 46.992 son PDF.

Seleccionamos 100 documentos al azar y realizamos un análisis de la calidad de los datos. Este procedimiento es el que se usará durante la realización del proyecto.

Estos 100 documentos tienen un total de 9.142 páginas y un total de 2.696.009 palabras.

#### 3.1 Metodología

En este proceso, cada documento se descompone en bloques de texto utilizando un sistema avanzado de OCR (Reconocimiento Óptico de Caracteres), que identifica y registra "bounding boxes" dentro del documento. Cada bloque se cataloga cuidadosamente con información precisa sobre su ubicación, como el número de página y las coordenadas en el documento, y se captura el texto interpretado dentro de esas coordenadas. Para esta tarea se emplea PyMuPDF, una herramienta robusta de Python diseñada para la extracción y manipulación eficiente de documentos PDF.

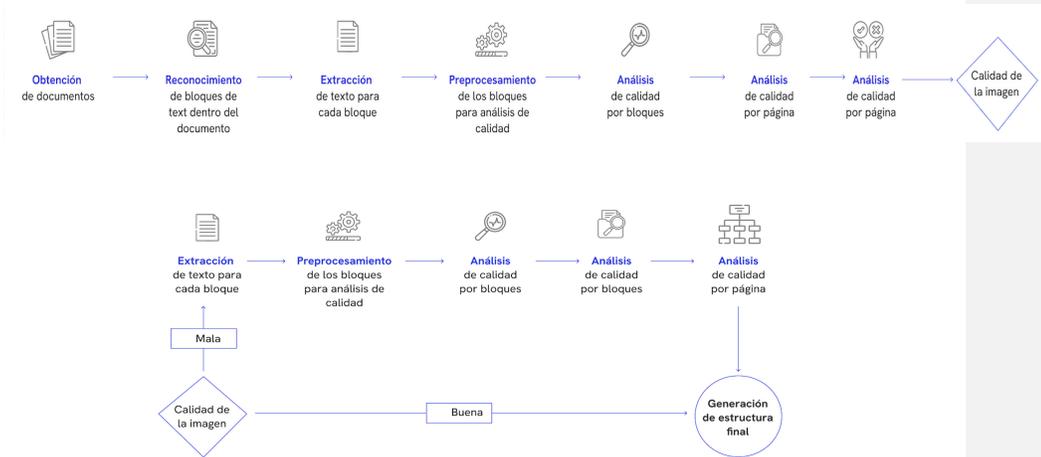
Luego se evalúa la calidad de la extracción, esto mediante la validación de cada palabra contra un diccionario de palabras reconocibles. Si la palabra existe entonces se contabiliza como válida. Luego se calcula el promedio de calidad de cada bloque, agrupándolo por página para luego obtener la calidad general del documento, lo que permite tener una evaluación detallada de la calidad del documento.

Los próximos pasos consideran poder aplicar un método que permita aplicar mejoras en caso de obtener una baja calidad de extracción, como, por ejemplo, utilizando un OCR de mejor calidad o un preprocesamiento más robusto del texto para evaluar su calidad.

Finalmente se crea un archivo por documento donde se registra el texto extraído, su ubicación respecto al archivo original, palabras reconocidas, palabras no reconocidas y el tiempo de procesamiento.

**Comentado [CH21]:** Favor, detallar la tecnología usada para esta acción.

### 3.2 Diagrama del proceso



En el proceso de extracción de datos, se comienza adquiriendo los documentos de interés. Luego, mediante automatización, se identifica la posición de los bloques de texto en cada documento.

Posteriormente pasan por el OCR (extracción de texto para cada bloque). Los datos son pre-procesados para evaluar la calidad de la extracción.

En el siguiente paso, se analizan las métricas de calidad que fueron obtenidas. Si estas métricas son satisfactorias, se genera la estructura final de datos inmediatamente.

En caso contrario, se recurre a un OCR avanzado para mejorar la extracción de texto, y finalmente, los datos extraídos con calidad mejorada se almacenan en la estructura final.

### 3.3 Definición de calidad

Para evaluar la calidad del texto lo que se hace es comparar cada palabra que se extrae con un diccionario de palabras que tiene el OCR, si la palabra aparece, la cuenta como correcta. Si la palabra no aparece en el diccionario, significa que está mal escrita, entonces la cuenta en el total de palabras. El indicador de calidad es:

$$\text{Calidad de oración} = \text{Palabras correctas} / \text{Palabras totales} \times 100$$

Las palabras que no se leen porque el OCR no las detecta no son parte del indicador de calidad de los datos en esta primera versión.

Aquí tenemos un ejemplo que ilustra el problema:

"Hola soy Alonso {palabra no reconocible}"

En este ejemplo hay 3 palabras que lee el documento de un total de 4 palabras, la calidad de la oración es 75%.

### 3.4 Ejemplo de segmentos en una página de un documento

1302 mil trescientos dos

MINISTRO EN VISITA EXTRAORDINARIA  
SR. CARLOS FARIAS  
(Adm. Nat. N° 155-2011)

OFICINA DE PARTES E INFORMACIONES  
30 ENE 2022

REMA CORTE DE APELACIONES DE SAN MARCO

REMANENCIA CUARENTA ALACOR  
30 ENE 2022

REMA CORTE DE APELACIONES DE SAN MARCO

SOLICITA COPIAS QUE INDICA

SR. MINISTRO EN VISITA EXTRAORDINARIA

GABRIEL BUENO LEAL, abogado por el Programa de Derechos Humanos del Ministerio de Justicia y Derechos Humanos, en autos ROL N° 155-2011 sobre delitos en perjuicio de MARIO STAFFLING, a US. respetuosamente digo:

Que, con el objeto de conocer las actuaciones realizadas en la causa y así poder colaborar en la presente investigación, vengo en solicitar copias, en formato digital, de estos autos;

POR TANTO,  
AL SEÑOR MINISTRO RUEGO, acceder a lo pedido;

*Gab. B.*  
15.028.164-6

Paulina Zamorano Valenzuela (S), Jefe Unidad Programa de Derechos Humanos de la Subsecretaría de Derechos Humanos del Ministerio de Justicia y Derechos Humanos, declara que este archivo es copia fiel del documento que se encuentra en el archivo físico de esta unidad (Fecha: 2021/02/02)

Lo primero es detectar los rectángulos que delimitan y rodean áreas de texto en el documento. Este procedimiento permite aislar segmentos específicos, mejorando así la precisión del reconocimiento de texto, pudiendo determinar la calidad de manera aislada en cada sección específica. Este método también permite detectar la posición efectiva de cada fragmento de texto en contenido de la página.

Ejemplo de archivo de datos respecto a la página.

La transcripción (OCR) que se está usando es una librería *open source* de Adobe. Esta genera un archivo en formato JSON que representa la estructura completa del documento dividido por páginas. Cada página del archivo JSON está compuesta por contenedores de texto con su correspondiente transcripción y la ubicación exacta de la caja en la página. Junto a cada transcripción, se presenta una métrica de calidad.

The image shows a JSON file on the left and a scanned document page on the right. Blue arrows indicate the mapping between the JSON structure and the document content:

- The JSON entry for "MINISTRO DE VISITA EXTRAORDINARIA" points to the header of the document.
- The JSON entry for "Sr. CARLOS PARIAS" points to the name of the person being visited.
- The JSON entry for "Oficina de partes" points to the stamp of the Office of Parts.

The document page contains the following text:

1302 mil trescientos dos

MINISTRO DE VISITA EXTRAORDINARIA  
Sr. CARLOS PARIAS  
Autos Rol N° 155-2011

OFICINA DE PARTES DE RECURSOS-COMUNES  
10 DE ENERO 2012  
EN LA CORTE DE APELACIONES DE SAN PABLO

RECORRIDO DE PARTES  
ALABORADO  
10 DE ENERO 2012  
EN LA CORTE DE APELACIONES DE SAN PABLO

SOLICITA COPIAS QUE INDICA

Sr. MINISTRO EN VISITA EXTRAORDINARIA

GABRIEL BUENO LEAL, abogado por el Programa de Derechos Humanos del Ministerio de Justicia y Derechos Humanos, en autos ROL N° 155-2011 sobre delitos en perjuicio de PABLO STICHELINI, a US, respectivamente digital

Que, con el objeto de conocer las actuaciones realizadas en la causa y así poder colaborar en la presente investigación, venga en solicitar copias, en formato digital, de estos autos.

POR TANTO:  
AL SEÑOR MINISTRO BUENO, acudir a lo pedido.

*[Signature]*  
15.01.2012

En este ejemplo el texto confundió Sr Carlos Parias con Carlos Farias y ahí la calidad fue 66.6%. En el bloque posterior mezcló la lectura de dos bloques Autos Rol 155-2011 lo unió con el timbre de oficina de partes. Aquí la calidad fue de 85.7%, no leyó mal sino mezcló dos textos que debieron ir separados, por eso la calidad del texto es más alta.



A continuación, se muestran más ejemplos de análisis de calidad para frases encontradas en el documento.

Documento	Texto	Porcentaje de Calidad
8-00816-6a_signed (1502-1841).pdf	.W; :.fi: "" \ -Σ""1""J	0%
-01772-1a (151-507)_signed.pdf	DICTADO POR DON MARIQ CAOZA ESPINOSA, MINISTRj EN VISITA EXTRAORDINARIA.	70%
-02118-3a (1302-1381).pdf	Identificación completa de los siguientes oficiales que prestaban servicios	100%

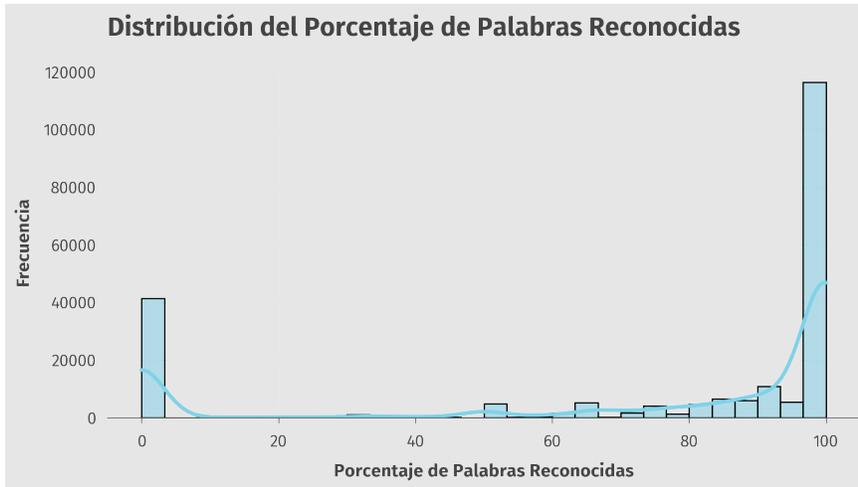
### 3.5 Estadísticas para 100 documentos

A continuación, se muestran las estadísticas para el análisis de calidad de los bloques en 100 textos escogidos al azar, los cuales tiene un total de 211.804 bloques. Hay que recordar que un 100% de calidad significa que el texto pudo ser reconocido completamente y un 0% de calidad indica que el bloque dentro del texto es ilegible.

Intervalo de calidad (Porcentaje de palabras reconocidas)	Cantidad de bloques dentro del intervalo
Menos de 40%	43.016
Entre 40% y 80%	22.923
Más de 80%	145.865

La comparación de las palabras se hace respecto al diccionario de palabras que contienen los OCR. En una segunda fase complementaremos ese diccionario con palabras propias del proceso, introduciendo las que más se repiten como ILTMA en el ejemplo de la sección 5.3.

A continuación, se muestra un histograma, que indica la cantidad de bloques que tienen un porcentaje de calidad determinado.



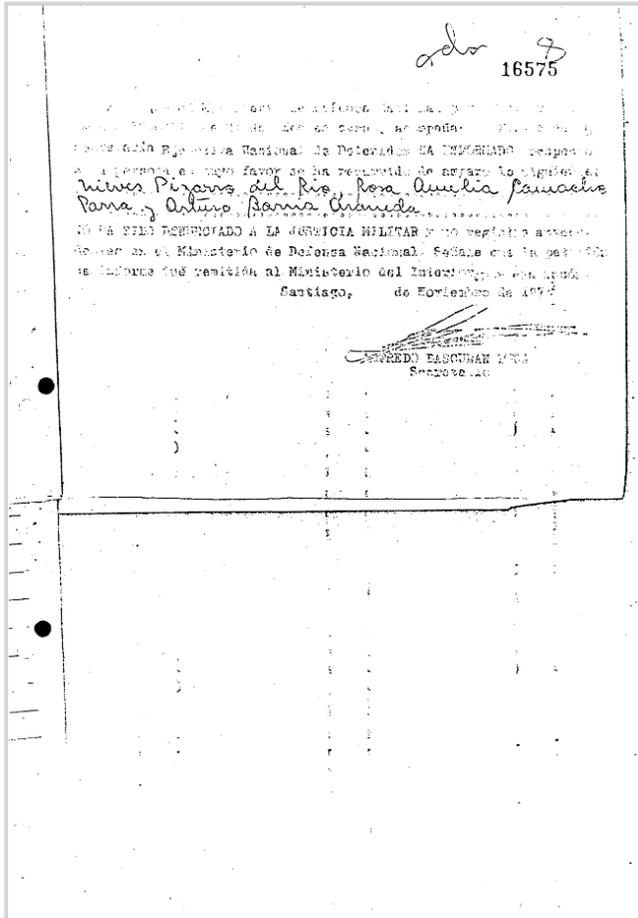
Este histograma muestra de manera más nítida lo que vimos en la tabla anterior. Existen dos extremos, bloque donde la calidad de las transcripciones es muy alta y bloques donde la calidad del texto transcrito es muy baja, después existe un conjunto de bloques intermedio donde la calidad varia, pero esos bloques representan un 10,8% de los bloques, mientras que los bloques con muy mala calidad representan un 20,3% de los bloques.

### 3.6 Tiempo de procesamiento y costo

Para los 100 documentos analizados, el tiempo de procesamiento fue de 30 minutos. Esto implica que para todos los documentos se requieren 10 días de procesamiento y cálculo. Esta es una primera estimación. En el trabajo identificamos varios métodos como puntos de mejora para hacer más eficiente el procesamiento de los documentos.

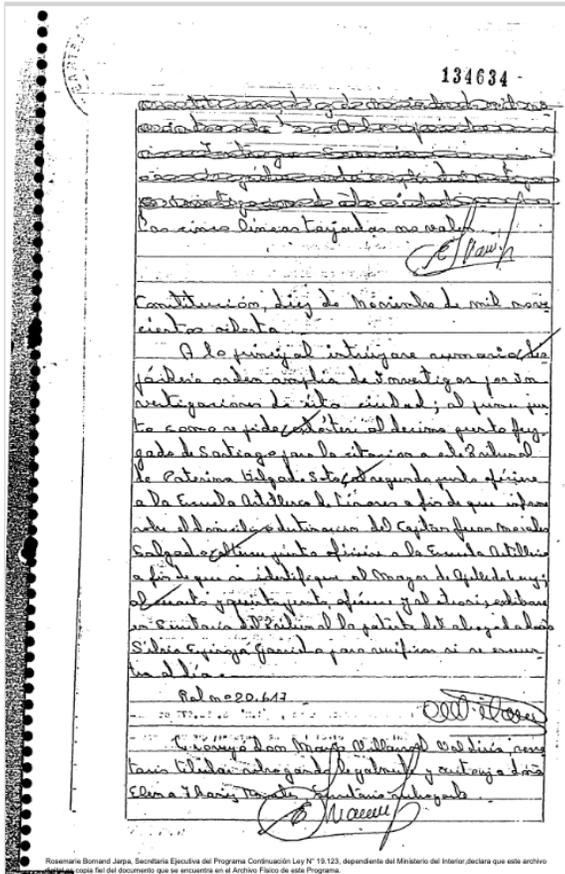
Respecto al costo, por el momento el método probado no tiene costo, sin embargo, para las páginas con baja calidad vamos a probar OCR alternativos pagados para ver si los identifica mejor.





En este ejemplo que también tiene un nivel de calidad de la transcripción bajo, hay anotaciones a mano en medio del texto general, después aparece una forma con un nombre que esta difícil de leer.<sup>3</sup> La calidad de esta página es 26.7%.

<sup>3</sup> Documento: 1-00081-1a (sin foliación).pdf (página 14)



Finalmente mostramos un texto escrito totalmente a mano.<sup>4</sup> La calidad de la transcripción de esta página es de 37.2%.

### 3.8 Próximos pasos

Para la siguiente versión, se van a trabajar en los siguientes elementos:

- Analizar la calidad de todos los documentos PDF con la versión no pagada de los OCR.
- Trabajar en la calidad de los datos de los archivos .tif.

<sup>4</sup> Documento: 8-00816-6a\_signed (1502-1841).pdf (página 11)

- Construir una distribución de las palabras que más se repiten que no están en el diccionario y que podrían servir para mejorar la precisión en la medición de la calidad de los documentos. El ejemplo más claro hasta ahora con los ejemplos mencionados es la palabra “ILTMA”
- Hacer una primera detección de nombres usando un diccionario de nombres y el padrón electoral y verificar su calidad, como “Carlos Parias” en vez de “Carlos Farias”.
- Reconocer imágenes en páginas y reconocer texto a mano en páginas y determinar la calidad de esos documentos.

## 4 Propuesta de módulos, sistemas y componentes a construir

Un módulo es un conjunto de funcionalidades dentro de un mismo dominio de acción que puede satisfacer requerimientos de manera focalizada en un stakeholder como también de manera transversal a ellos.

### 4.1 Módulo de documentos

Su principal objetivo es el manejo del ciclo de vida de un documento. Las características de este módulo son:

- **Inmutabilidad de los documentos:** No permite modificaciones a los archivos originales
- **Versionado:** Permite agregar enmiendas a un archivo, pero registrando versiones explícito que sea parte del modelo de referencia.
- **Referenciación:** Provee un identificador único, inmutable, para cada documento del repositorio, a usarse a lo largo de la plataforma.
- **Metadata archivística:** Mediante el uso de metadata soporta la norma archivística ISAD(G)
- **Trazabilidad:** Toda operación es relacionada a un individuo, un instante en el tiempo y una IP.

### 4.2 Módulo de procesamiento

Extrae contenidos utilizables en los procesos de investigación desde documentos de la plataforma.

- **OCR:** Implementa múltiples soluciones de OCR permitiendo utilizar los más adecuados a cada tipo de fuente (escrito a máquina, manuscrito, formularios, párrafos, distintos idiomas).

- **Imágenes:** Extrae imágenes y figuras individuales desde documentos o archivos con múltiples especímenes.
- **Transcripción:** Transcripción (a texto) de voces habladas.
- **Coordenadas:** Genera coordenadas específicas para cada elemento de contenido (documento, página, párrafo, ubicación en la página).
- **Indexación:** Indexa contenidos de manera de permitir búsquedas eficientes
- **Categorización:** Atribución de categorías para documentos y elementos de contenido individuales.
- **Curatoría y Mantenimiento:** Soporta correcciones/enmiendas manuales a los contenidos.
- **Trazabilidad:** Registro detallado e inviolable de toda actividad (automatizada o manual) realizada sobre los contenidos.

### 4.3 Módulo de investigación

Flujos y actividades para investigadores relacionadas con el ciclo de vida de las hipótesis.

- **Construcción de Hipótesis:** Autoría de hipótesis como una colección de eventos y relaciones entre entidades que ocurren en el tiempo.
- **Selección de Contenidos y Entidades:** Exploración, búsqueda y descubrimiento, manual y asistido, de entre contenidos y entidades de la plataforma. Búsquedas guardadas que permiten notificación ante nuevas coincidencias aportadas a la plataforma.
- **Historial:** Mantenimiento de notas y registro de actividades relacionadas con la hipótesis, además de modelamiento de estados formales (Abierto, En Curso, Cerrado, Confirmado, etc.)
- **Visualización:** Organización de la información relacionada a una hipótesis usando grafos, tablas, líneas de tiempo y otras presentaciones.

## 4.4 Módulo de anotaciones

Mantenimiento de entidades a utilizar durante el proceso de investigación.

- **Repositorio de Entidades:** Almacenamiento, edición y manejo de versiones para:
  - **Personas:** víctimas, testigos, agentes, osamentas
  - **Lugares:** recintos, lugar de inhumación, sitios de interés
  - **Organizaciones:** políticas, religiosas, de Inteligencia, Fuerzas Armadas, entre otras
  - **Relaciones:** familiares, captura, retención, traslado, tortura, asesinato, entre otras
  - **Eventos:** captura, retención, traslado, tortura, asesinato, entre otras.
  - **Periodos:** fechas, rangos de fechas.
- **Modelamiento de atribuciones aproximadas:**
  - Una persona de entre varias posibles.
  - Períodos aproximados.
  - Lugares aproximados.
  - Eventos con información incompleta.
- **Extracción automatizada:** Creación de entidades desde contenidos, con opciones de revisión manual.
- **Anotación manual y automatizada:** Vinculación de contenidos con entidades vía modelos y vía actividad manual.
- **Integración con fuentes externas:** API para importar anotaciones desde herramientas externas o como importación masiva.

## 4.5 Módulo de trayectorias

- **Manejo de Múltiples hipótesis:** Generación de múltiples hipótesis de trayectorias de víctimas.
- **Identificación de entidades:** Dentro de una trayectoria, de existir, debe poder identificarse uno o más: personas: víctimas, testigos, agentes, osamentas y lugares: recintos, lugar de inhumación, sitios de interés.
- **Persistencia:** Guardar una trayectoria para retomarla a posteridad.

## 4.6 Módulo de familiares

- **Datos familiares:** Disponibilidad de una base de datos de los familiares con información de contacto.
- **Seguimiento:** Revisión de los trámites y consultas digitales solicitadas.
- **Bitácora:** Seguimiento de los acompañamientos y mediaciones realizadas.
- **Encuentros:** Gestión de los encuentros nacionales y mesas de trabajo.
- **Financiamiento:** Gestión de financiamiento para familiares.

- **Reporteías familiares:** Comunicaciones automatizadas vía correo y newsletter.
- **Enriquecimiento:** Robustecer las mallas familiares con información emergente (agregar datos de familiares nuevos)
- **Registro civil:** Integración del registro civil para obtención de datos de familiares

#### 4.7 Terreno

- **Documentos trabajos:** Ingreso y visualización de documentación generada por los trabajos en terreno.
- **Financiamientos familiares:** Planificación de recursos necesarios para el trabajo y la participación de los familiares.

#### 4.8 Jurídica

- **Integraciones posibles:** Revisan datos en la página del poder judicial
- **Trazabilidad:** Identificar una causa por cualquiera de sus roles, los cuales van cambiando conforme avanza el proceso.
- **Estado actual:** Conocer el ultimo estado procesal de una causa y enlazar toda su información con los documentos existentes con lo cual se tiene trazabilidad de la causa.
- **Investigación:** Poder realizar búsquedas y filtros sobre las causas.
- **Alertas:** Tener una comunicación con el área social respecto al avances de las causas generando un sistema de alertas.

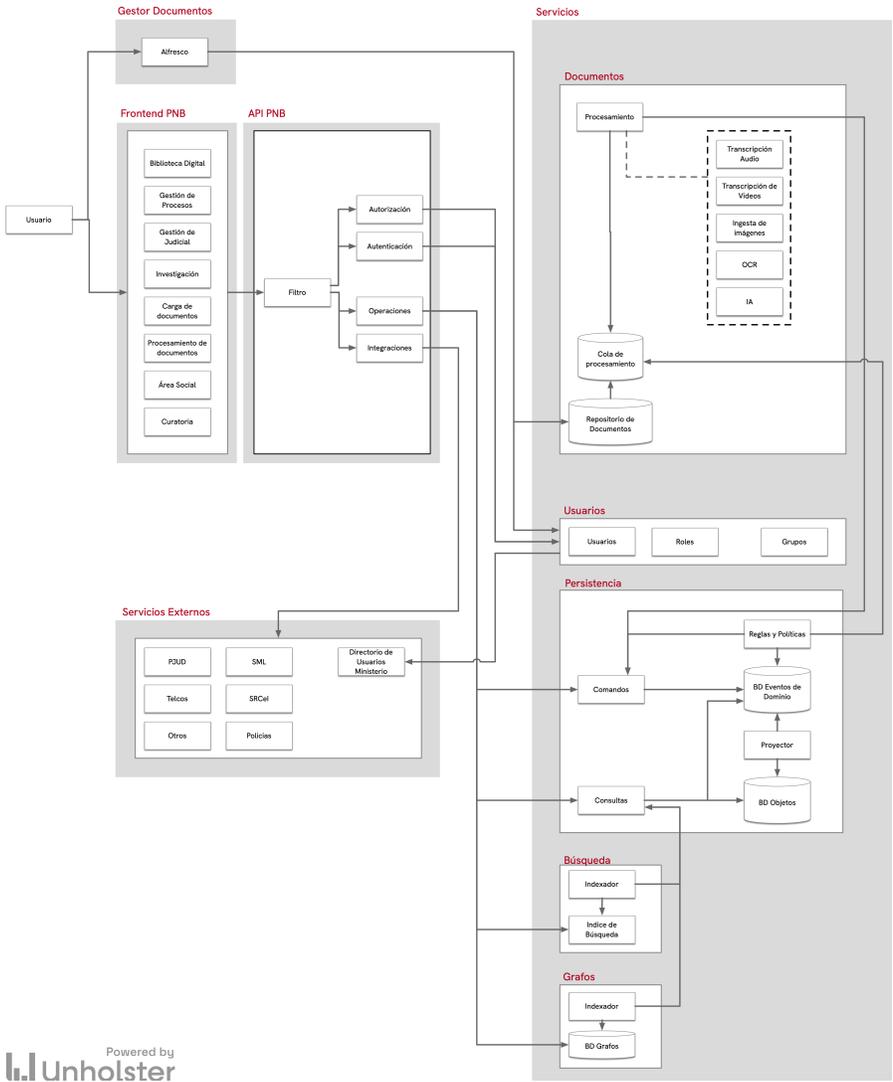
#### 4.9 Servicio Médico Legal

Conexión con sus bases de datos para obtener información sobre:

- Estatus de víctimas.
- Resoluciones judiciales.
- Toma de muestras (donadores de ADN).
- Datos antropomórficos.
- Lugares de interés.
- Actividades periciales.
- Inventario osamentas.

# 5 | Arquitectura final

## Plan Nacional de Búsqueda - Arquitectura General



**Comentado [CH22]:** Por favor, explicar el esquema.

Comentario Unholster: Está explicado en el punto "COMPONENTE" de mas abajo.

## 5.1 Principales Características

### 5.1.1 Orientada al Evento

La arquitectura propuesta se basa en el principio de que la operación del sistema es una secuencia de comandos que puede ejecutar un usuario dependiendo su dominio de acción, lo cuales generan eventos que modifican el estado de las entidades u objetos de negocio de los que tiene conocimiento el sistema completo.

Esto significa que se registran en una base consolidada todas las acciones que ejecutan los usuarios en su dominio de negocio y que afectan el estado de los diferentes objetos o entidades de las que tiene conocimiento el sistema. Esto garantiza poder reconstruir el estado del sistema en cualquier momento del tiempo, así como también de cada objeto o entidad de manera aislada, lo que garantiza una capacidad de trazabilidad y auditoria granular de cualquier entidad.

### 5.1.2 Interoperabilidad

La arquitectura propuesta por Unholster considera mecanismos para recibir y entregar datos desde y hacia entidades externas.

### 5.1.3 Escalabilidad

El diseño propuesto permite escalar horizontalmente en los servicios que implementan la lógica de negocios de cara a los usuarios finales

### 5.1.4 Licencias

Los sistemas operativos considerados, librerías y dependencias y tecnologías de base de datos a utilizar son libres de licencias, de uso libre y código abierto.

Para servicios de monitoreo y alerta, así como algunas soluciones OCR a incorporar se consideran servicios libres de licencia, pero pagados como SaaS.

## 5.2 Componentes

### 5.2.1 Gestor de Documentos

Proveerá la interfaz de usuario y todas las funcionalidades relacionadas con la colección digital de contenido. Es un paquete de software del tipo CMS (Content Manager System) que permite implementar sobre las mejores prácticas de almacenamiento de contenido, clasificación y búsqueda de documentos y archivos de toda índole.

Además, estará integrado con un backend de documentos que proveerá de todas las funcionalidades para la transcripción y procesamiento del contenido desde los formatos originales hacia la transformación a datos en texto manejables por el sistema y los usuarios.

### 5.2.2 Frontend PNB

Pone disposición de los distintos usuarios las funcionalidades que requiere para realizar su trabajo, de acuerdo con la experiencia de usuario acordada. En otras palabras, es la cara visible del sistema de cara a todos los usuarios.

Cada aplicación cuenta con su frontend particular que se conecta a la API PNB que controla las acciones posibles de realizar en el contexto de cada aplicación, dependiendo de los requerimientos funcionales que realicen los usuarios en etapas futuras de desarrollo del proyecto.

### 5.2.3 API PNB

Resuelve todos los requerimientos de la capa de frontend en cuanto a materializar las acciones que realiza el usuario a través de la UI.

Controla el acceso de los usuarios y los permisos de estos para realizar las diferentes tareas.

La API como parte de la capa de aplicación se encarga además de controlar el acceso de los usuarios y los permisos a cada una de las acciones posible de realizar, garantizando de esta forma que no sea posible ejecutar acciones para las cuales no se está autorizado.

### 5.2.4 Servicios Externos

La arquitectura contempla la integración con servicios externos. Cada integración será realizada con la mejor tecnología disponible en cada caso y servicio a integrar. Lo que depende fundamentalmente de cómo cada administrador de un sistema externo disponga de la modalidad a la cual se accederá a la información de interés para el PNB.

**Comentado [CH23]:** Se trata de un archivo que se distingue de una biblioteca pues en esta última se alojan libros y en nuestro archivo diversos tipos documentales.

Respuesta Unholster: Si bien el termino "Biblioteca Digital" cubre el concepto deseado, lo cambiamos por un concepto mas genérico como es "Colección Digital de Contenido".

En el numeral 7 se describen los servicios externos con los cuales, eventualmente, se integrará la plataforma.

### 5.2.5 Documentos

Es el backend del gestor documental (CMS) y está a cargo de recibir los archivos en los diferentes formatos, contener la meta data<sup>5</sup> y conocer la ubicación de cada uno de ellos en sus diferentes versiones y etapas del flujo de proceso. Además, proveerá de todas las funciones relativas a los procesos archivísticos necesarios y de etiquetado que se requieren para procesar los archivos que podrían contener información valiosa para el PNB.

### 5.2.6 Usuarios

La componente de usuarios se encarga de administrar y gestionar los permisos y accesos de acuerdo con las reglas que se designen para los diferentes roles/grupos. Adicionalmente tiene la misión de gestionar la conexión a directorios de usuarios de servicios externos que requieran acceso.

### 5.2.7 Persistencia

La componente de persistencia se encarga de almacenar toda la información relativa a los objetos de negocio que manejará el sistema, además de todos los eventos en los que participe dicha entidad y los estados que tomará durante la operación del sistema. En resumen, es la base de datos principal del sistema en donde se almacenará toda la información que puede ser utilizada por los diferentes módulos.

### 5.2.8 Búsqueda

La componente de Búsqueda proveerá de la información para encontrar de manera eficiente los estados de las diferentes entidades que manejará el sistema, esto es relevante debido a que durante su operación se espera que el sistema almacene una importante cantidad de información que debe ser fácil de encontrar y procesar por los diferentes módulos que verán los usuarios.

---

<sup>5</sup> Entendemos por metadatos como cualquier información que no pertenece al documento pero si lo describe y/o caracteriza. Son datos sobre los datos que contiene el documento.

**Comentado [CH24]:** Qué están entendiendo por metadatos?

Respuesta Uholster: Entendemos por metadatos como cualquier información que no pertenece al documento pero si lo describe y/o caracteriza. Son datos sobre los datos que contiene el documento.

**Comentado [CH25]:** "funciones relativas a los procesos archivísticos necesarios"

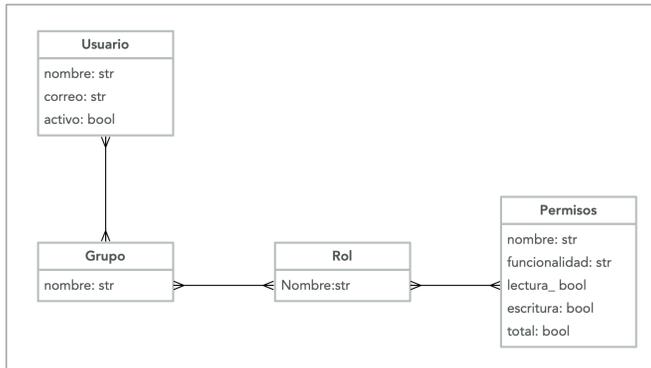
### 5.2.9 Grafos

La componente de grafos almacenará las relaciones que se establezcan entre la información capturada y curada desde los diferentes documentos y archivos, con las cuales se dará forma a la ontología que permiten realizar las hipótesis, investigaciones y conclusiones de cada caso.

## 6 Definición base de datos

### 6.1 Modelo de Usuarios

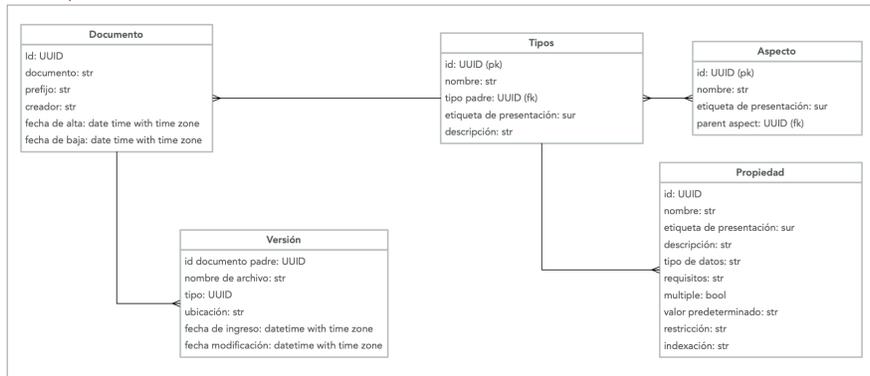
**Modelo de Usuarios**



En este modelo de datos, se establece una estructura jerárquica para gestionar la autorización y los permisos en un sistema. Los usuarios son la entidad principal y pueden pertenecer a uno o varios grupos. Cada grupo tiene asignados diversos roles, y cada rol define conjuntos específicos de permisos. De esta manera, los usuarios heredan los permisos a través de la estructura de grupos y roles a los que pertenecen. Esto proporciona una forma eficiente y escalable de administrar la autorización en el sistema, ya que los cambios en los permisos se pueden realizar a nivel de roles y grupos, afectando automáticamente a todos los usuarios asociados. Además, esta estructura facilita la gestión de permisos al evitar la necesidad de asignar permisos individualmente a cada usuario, permitiendo una administración más centralizada y coherente.

## 6.2 Modelo Repositorio

Modelo Repositorio



En primer lugar, se introduce un modelo de Documento que almacena información descriptiva sobre las condiciones de ingreso de documentos. A su vez, el modelo de Versión gestiona las múltiples versiones de estos documentos, permitiendo un seguimiento preciso de los cambios a lo largo del tiempo. Cada documento se encuentra asociado a un modelo Tipo, que actúa como una plantilla, definiendo las características esenciales para su clasificación según la norma archivística ISAD(G) y proporcionando atributos para describir su contenido de manera flexible o rigurosa según sea necesario.

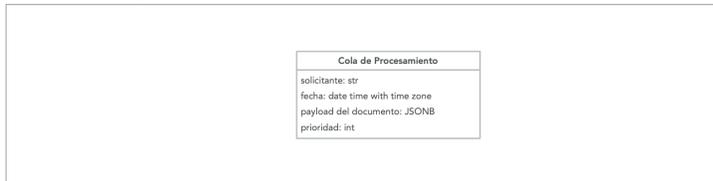
La instancia del modelo Tipo se enriquece mediante relaciones con el modelo Propiedad, que especifica los atributos, ya sean metadatos o no, necesarios para la plantilla. Estos atributos son esenciales para adecuar la documentación a las normativas archivísticas y brindan la posibilidad de describir con detalle el contenido del documento.

Por último, se ha incorporado el modelo Aspecto, similar a Propiedad, pero con un enfoque en funcionalidades preconstruidas. Por ejemplo, el aspecto de geolocalización facilitará la inclusión de coordenadas en un mapa al ingresar un documento asociado a un tipo que tenga el aspecto "geolocalizable" activado, proporcionando una experiencia de usuario más intuitiva y flexible según las necesidades específicas de cada caso.

Este modelo se presenta como una primera versión, sujeta a iteraciones continuas a medida que avanzamos en el proyecto, asegurando una evolución constante y adaptativa a los requerimientos emergentes.

### 6.3 Modelo cola de procesamiento

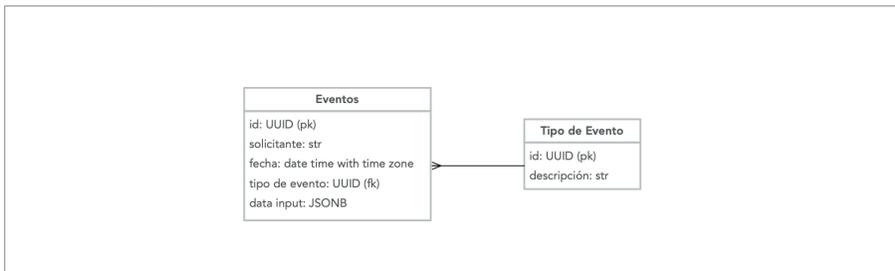
Cola de Procesamiento Documentos



En el ámbito de la carga de documentos, especialmente al enfrentarse a cargas masivas, resulta imperativo implementar un sistema de cola para gestionar eficientemente el tiempo de procesamiento asociado a cada documento. Esta estructura de cola garantiza un manejo ordenado y secuencial de los documentos, permitiendo que el sistema los procese uno a uno en función de su llegada. Un aspecto destacado en este proceso es la inclusión de un atributo de prioridad, el cual confiere la capacidad de procesar de manera inmediata aquellos documentos ingresados de forma individual con prioridad más alta, antes de abordar los documentos restantes en la cola. Este enfoque estratégico no solo optimiza el rendimiento del sistema, sino que también asegura la pronta atención de documentos críticos o urgentes, mejorando así la eficiencia global del proceso de carga.

### 6.4 Modelo eventos de dominio

Eventos del Dominio



En una arquitectura basada en eventos, cada interacción en el sistema queda meticulosamente registrada en una tabla de eventos. Esta tabla actúa como un repositorio fundamental, almacenando la meta data asociada con cada evento específico, el tipo de evento que se va a ejecutar y la información necesaria para llevar a cabo la acción correspondiente. Cada evento, al ser desencadenado, sigue un curso de acción determinado por las reglas de negocio predefinidas. Esta estructura basada en eventos no solo proporciona una visión detallada de todas las actividades en el sistema, sino que también destaca la importancia de la auditabilidad del sistema. La capacidad de rastrear y revisar de manera sistemática cada evento y entidad en el tiempo permite una mayor

transparencia y trazabilidad en las operaciones del sistema, lo que se traduce en un nivel más alto de confianza y seguridad en el entorno tecnológico.

## 6.5 Modelo de objetos

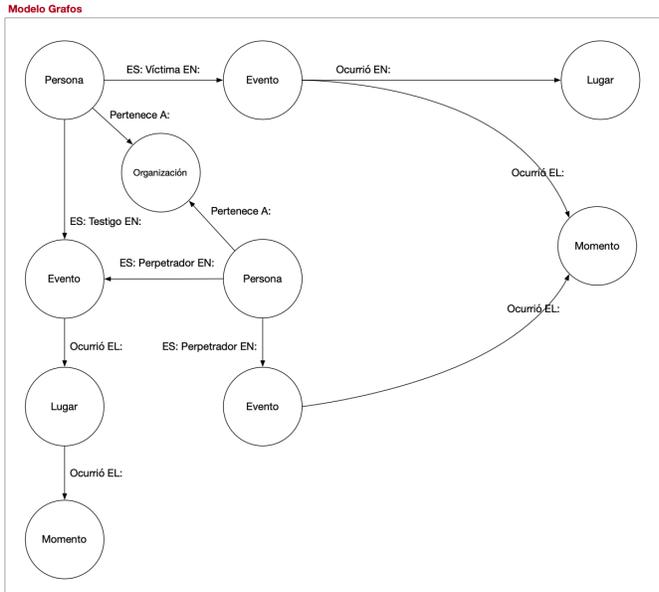


La base de datos de objetos almacena objetos resueltos originados a partir de los eventos de dominio acumulados en el sistema.

A medida que se reciben nuevos eventos el proyecto actualiza los objetos almacenados.

La base de objetos sirve de fuente para poblar los índices y grafos del sistema.

## 6.6 Modelo de grafo



El modelo de base de datos que empleamos utiliza un enfoque de grafos para gestionar la información, con clases de entidades que incluyen Persona, Organización, Evento, Lugar y Momento. Cada clase tiene atributos específicos que capturan sus propias características distintivas. Además, se incorporan subclases de entidades, cada una con atributos más detallados y especializados. Ejemplo de esto son las subclases para Persona: Perpetrador, Víctima, Buscador o Familiar.

Para estructurar las relaciones entre estas entidades, seguimos una ontología basada en el sistema Angelus. Esta ontología define y organiza las múltiples relaciones posibles entre las entidades, proporcionando un marco conceptual sólido para comprender la red de datos. Estas relaciones, permiten un análisis significativo y detallado de la interconexión entre las entidades en el sistema de base de datos de grafos.

Pese a que la ontología aún se encuentra en definición, a continuación, se presentan algunas relaciones a considerar:

Entidad	Relación	Entidad
Persona	<ul style="list-style-type: none"> <li>• Es la misma que...</li> <li>• Es buscado por...</li> <li>• Es encontrado por...</li> <li>• Es subordinado de...</li> <li>• Es representado por...</li> <li>• Es familia de...</li> </ul>	Persona
Persona	<ul style="list-style-type: none"> <li>• Es víctima de...</li> <li>• Es perpetrador de....</li> <li>• Ordenó evento...</li> <li>• Es responsable de...</li> <li>• Es denunciante de...</li> </ul>	Evento
Persona	<ul style="list-style-type: none"> <li>• Es miembro de organización...</li> </ul>	Organización
Persona	<ul style="list-style-type: none"> <li>• Está en....</li> </ul>	Lugar
Evento	<ul style="list-style-type: none"> <li>• Tiene lugar en...</li> <li>• El inicio del evento ocurrió en...</li> <li>• El final del evento ocurrió en....</li> </ul>	Lugar
Lugar	<ul style="list-style-type: none"> <li>• Es sede de...</li> </ul>	Organización
Lugar	<ul style="list-style-type: none"> <li>• Es el mismo que...</li> <li>• Es parte de...</li> </ul>	Lugar
Evento	<ul style="list-style-type: none"> <li>• Es el mismo que...</li> <li>• Es desencadenado por</li> </ul>	Evento

## 7 Interoperabilidad

La interoperabilidad es un aspecto fundamental en el diseño y construcción del sistema. A la fecha hemos identificado varios potenciales proveedores de información relevante con los que necesitamos establecer integraciones efectivas.

### 7.1 Sistemas Internos

#### 7.1.1 Alexandria

Se planea una migración de datos, considerando cambiar desde la norma MARC21 a ISAD(G). Para llevar a cabo esta migración, se evaluará la utilización de la API proporcionada por Alexandria. De esta manera se realizará una extracción de la información de forma más directa, limpia y eficiente. Eventualmente el objetivo es dejar de utilizar esta plataforma, ya sea reemplazándola por otra opensource, que nos permita integración con la plataforma a desarrollar o reemplazándola por desarrollo propio.

#### 7.1.2 Sistema DDHH

Otro objetivo considera migrar los datos del Sistema de Derechos Humanos. Sin embargo, hasta el momento contamos con poca documentación del sistema, lo que plantea un desafío para obtener los datos. Dado lo anterior, se propone usar un proceso de web scraping sobre la plataforma para obtener y transferir datos con mayor seguridad.

### 7.2 Sistemas Externos

#### 7.2.1 SRCEI – Servicio de Registro Civil e Identificación

Utilizaremos una API construida a medida por la contraparte y se accederá al sistema "monito" para consultas manuales. Un desafío en este proceso es que hay cerca de 500 personas sin rutificar, a las cuales legalmente no se puede asignar un RUT, lo que nos lleva a buscar un enfoque alternativo y manual para obtener su información que nos permita identificar estas personas mediante otra llave.

#### 7.2.2 SML – Servicio Médico Legal

El principal desafío de esta integración es que la mayor parte de la información con la que cuenta el Servicio Médico Legal está en formato físico (no digital). Sin embargo, tenemos información de que está en proceso de digitalizar documentos, en particular, aquellos relacionados

con trabajos en terreno, datos de osamentas y análisis forenses. El método de integración con ellos se definirá cuando se cuente con mayor avance en su proceso de digitalización.

### 7.2.3 PJUD – Poder Judicial como proveedor de información

Existe una oportunidad de integración con la página del Poder Judicial, que permita la obtención de información para el monitoreo de investigaciones y causas. Es importante señalar que las investigaciones y causas de interés no son de carácter público, por lo que se requiere autenticación de un abogado relacionado con ellas para poder revisar los datos. Esta integración dará al sistema la posibilidad de realizar un seguimiento directo de las actualizaciones en “tiempo real”.

**Comentado [CH26]:** Para el PNB esta posibilidad no supone un gran interés, para que sea considerada en razón del esfuerzo necesario para lograr esta funcionalidad.

Pregunta Unholster: ¿Con este comentario se espera que no se considere la integración con el PJUD o que no se considere como una oportunidad?

### 7.2.4 PDI - Policía de Investigaciones

La integración con Policía de Investigaciones es de utilidad para el Plan Nacional de Búsqueda, ya que permite acceder a información crucial sobre los ingresos y egresos del país. Sin embargo, en este momento, el método de integración con este sistema aún no se ha definido y requiere un análisis más detallado para determinar la estrategia más adecuada, debido a que se requiere de autorizaciones que no son fáciles de conseguir.

### 7.2.5 PJUD - Poder Judicial como usuario de información

En el caso del Poder Judicial como usuario de información, en lugar de una integración, lo que necesitamos es proporcionar acceso a la plataforma del Plan Nacional de Búsqueda mediante un perfil de usuario específico. El PJUD se encuentra actualmente en proceso de digitalización, lo que lo convierte en una fuente potencialmente valiosa de nuevos documentos para enriquecer el sistema.

**Comentado [CH27]:** Esto en realidad es Poder Judicial que está por iniciar un proceso de digitalización de expedientes.

## 8 Wireframes y flujo de navegación

La creación de wireframes y flujo de navegación se centró en las vistas prioritarias definidas en las reuniones de requerimiento, estas vistas ilustran la búsqueda de contenidos en la plataforma, resuelto en el marco de la V01, incorporando inicialmente funcionalidades esenciales que debe incorporar esta sección.

Inicialmente trabajamos un flujo descriptivo de navegación acompañado de wireframe en media fidelidad, que incluye login / buscador general / resultados y visualización de contenidos-documento. El resultado de esa exploración a continuación.



## 9 Definición de usuarios

En esta fase de baja madurez del diseño del sistema no tenemos aun información suficiente como para definir los tipos de usuario y la segregación de funcionalidades e información.

Creemos que esto emergerá espontáneamente en la medida que se avance con los diferentes módulos funcionales.

## 10 Maquetas

Producto de la exploración proveniente de los wireframes y flujo de navegación, desarrollamos en alta fidelidad las maquetas resultantes de ese ejercicio. A continuación, mostraremos las vistas junto a una descripción general de sus funcionalidades y características.

### Login de Usuario

Vista que permite acceder a la plataforma mediante combinación de usuario y contraseña.

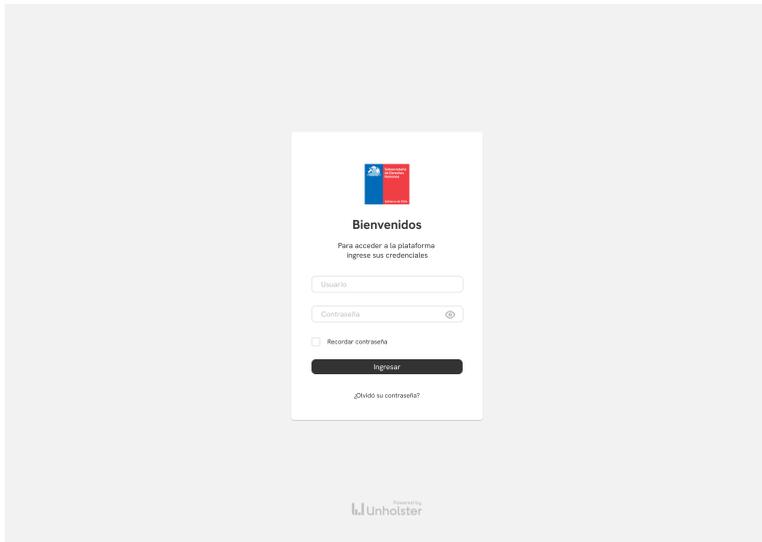


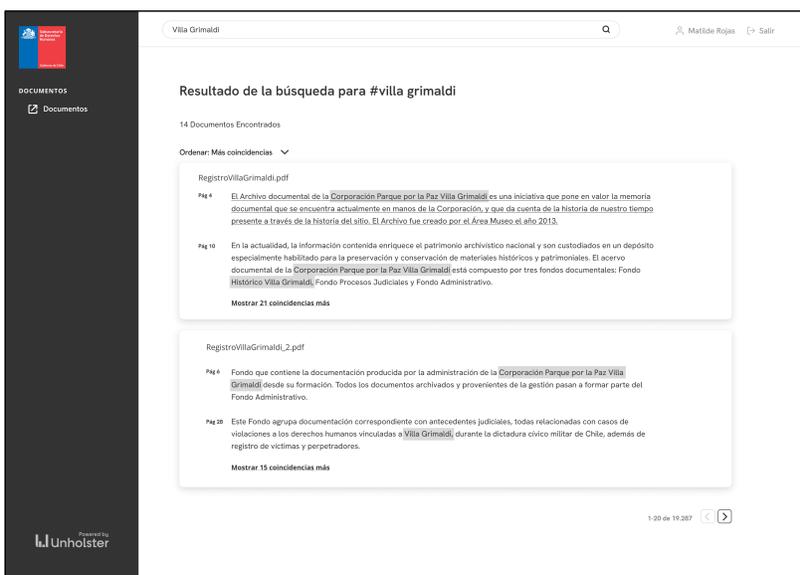
Ilustración 5. Vista login de usuario



## Búsqueda de Contenidos

A través de un buscador principal presente en el header de la plataforma, se podrá buscar contenidos presentes en los documentos almacenados en el sistema, puede ser mediante palabras claves o frases que permitan llegar a un resultado más exacto.

La plataforma desplegará todos los resultados presentes en la biblioteca de contenidos que incluyan la palabra o frase en cuestión, indicando el documento donde se encuentra, las páginas relacionadas y un extracto destacado del párrafo que lo contiene, permitiendo acceder desde los mismos párrafos a la página que se necesita consultar. Los documentos encontrados con coincidencias se podrán ordenar y navegar a través de paginación.



The screenshot shows a search interface with a search bar at the top containing 'Villa Grimaldi'. Below the search bar, the results are displayed under the heading 'Resultado de la búsqueda para #villa grimaldi'. It indicates '14 Documentos Encontrados' and 'Ordenar: Más coincidencias'. Two document entries are visible:

- RegistróVillaGrimaldi.pdf**
  - Párr 4**: El Archivo documental de la **Corporación Parque por la Paz Villa Grimaldi** es una iniciativa que pone en valor la memoria documental que se encuentra actualmente en manos de la Corporación, y que da cuenta de la historia de nuestro tiempo presente a través de la historia del sitio. El Archivo fue creado por el Área Museo el año 2013.
  - Párr 10**: En la actualidad, la información contenida enriquece el patrimonio archivístico nacional y son custodiados en un depósito especialmente habilitado para la preservación y conservación de materiales históricos y patrimoniales. El acervo documental de la **Corporación Parque por la Paz Villa Grimaldi** está compuesto por tres fondos documentales: Fondo Histórico Villa Grimaldi, Fondo Procesos Judiciales y Fondo Administrativo.
  - Mostrar 21 coincidencias más
- RegistróVillaGrimaldi\_2.pdf**
  - Párr 6**: Fondo que contiene la documentación producida por la administración de la **Corporación Parque por la Paz Villa Grimaldi** desde su formación. Todos los documentos archivados y provenientes de la gestión pasan a formar parte del Fondo Administrativo.
  - Párr 20**: Este Fondo agrupa documentación correspondiente con antecedentes judiciales, todas relacionadas con casos de violaciones a los derechos humanos vinculadas a **Villa Grimaldi**, durante la dictadura cívico militar de Chile, además de registro de víctimas y perpetradores.
  - Mostrar 15 coincidencias más

At the bottom right of the search results area, it shows '1/20 de 19.287' with navigation icons.

Ilustración 6. Vista búsqueda de contenidos

## Visualización de Contenidos-Documentos

Al hacer click en el párrafo que se desea consultar, se desplegará la vista de la página específica del documento, destacando las coincidencias de búsqueda, mostrando la descripción del documento, acceso al documento original. Constará también con la posibilidad de realizar una nueva búsqueda en el documento, mostrando cantidad de coincidencias e incorporando la funcionalidad de poder navegar y posicionarse sobre ellas.

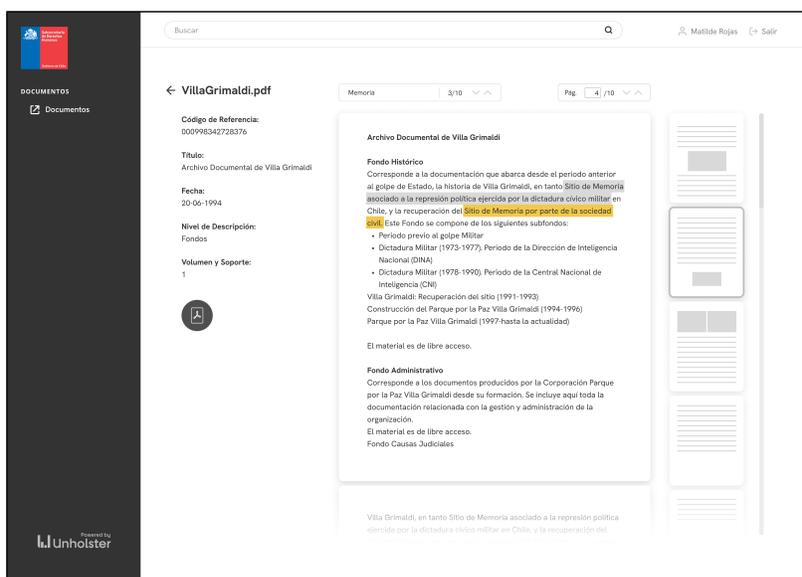


Ilustración 7. Vista visualización de contenidos-documentos